



Companion Proceedings of the 10th International Conference on Learning Analytics & Knowledge LAK20

Celebrating 10 years of LAK: *Shaping the future of the field*

March 23-27, 2020
The University of Frankfurt,
Frankfurt, Germany

Organized by:



TECHNISCHE
UNIVERSITÄT
DARMSTADT



This work is published under the terms of the Creative Commons Attribution- Noncommercial-ShareAlike 3.0 Australia Licence. Under this Licence you are free to:



Share — copy and redistribute the material in any medium or format

The licensor cannot revoke these freedoms as long as you follow the license terms.

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

Non-Commercial — You may not use the material for commercial purposes.

NoDerivatives — If you remix, transform, or build upon the material, you may not distribute the modified material.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

LAK 2020 Program Chairs' Welcome

We are pleased to welcome you to the *Tenth International Conference on Learning Analytics and Knowledge (LAK20)*, organized by the Society for Learning Analytics Research (SoLAR). This year's conference is hosted by the Goethe University in the beautiful city of Frankfurt, Germany on March 23–27, 2020, a place of tremendous importance and rich history of science, art, and philosophy. While this is the first time that LAK is organized in Germany, we want to acknowledge that it is a tremendous collaborative effort by the international community of learning analytics researchers and practitioners. Being the tenth anniversary of the conference, the theme for LAK20 is *"Celebrating 10 years of LAK: Shaping the future of the field"* and focuses on celebrating the achievements of the learning analytics community in the first ten years as well as on charting a pathway for the next ten years. The LAK20 conference is intended for both researchers and practitioners, and we invite them to come and join a proactive dialogue around the future of learning analytics and its practical adoption. We further extend our invite to educators, leaders, administrators, and government and industry professionals interested in the field of learning analytics and related disciplines.

Continuing the trend from previous LAK editions, we received record numbers of high-quality submissions. The research track received 261 submissions (137 full paper submissions and 124 short paper submissions), which represents a 23% increase in the total number of submissions. Overall, from the 261 research papers, the program committee worked very hard to select 80 papers (50 full research papers and 30 short research papers). As usual, the research papers are published as Proceedings by the ACM.

In addition to the full and short papers of the research track, we have accepted 10 practitioner reports, 41 research posters, 11 practitioner posters, 8 demos, 11 doctoral consortium submissions, and 22 workshop proposals. Some of these workshops had their own call for papers and their accepted submissions are included in these Companion Proceedings as well. We are most grateful for all the hard work by the practitioner, workshop, poster & demo and doctoral consortium chairs as well as by the program committee and their insightful and constructive comments and reviews. In total, over 1,000 reviews and meta-reviews have been provided by the LAK20 program committee. These Companion Proceedings could not have been done without their generous help and support.

While we celebrate the amazing achievements from the past 10 years of LAK, we are also very aware of the inherently interdisciplinary nature of the learning analytics community, with many different theoretical and methodological stances. The vision of SoLAR for the LAK conference is to build a welcoming and inclusive space where learning analytics researchers and practitioners with diverse viewpoints can engage in productive conversations that will only strengthen our impact and contribution to the improvement of teaching and learning for years to come.

Our hope is that the LAK20 participants and the readers of these proceedings will strengthen their connections with the rapidly growing learning analytics community. We believe that the use of data and analytics for the advancement of learning is only possible through productive dialogue between practitioners, researchers, policymakers, and industry professionals. In this regard, the LAK conference plays a central role in bringing those different stakeholders together to discuss the recent advancements in the learning analytics field and how they can be used for research-informed practice and policy development. We hope that these proceedings will provide a solid foundation for discussing our successes and failures in the past ten years, but more importantly the challenges and opportunities that lie ahead of us for the next ten years.

Vitomir Kovanović
University of South Australia
Australia

Maren Scheffel
Open Universiteit
The Netherlands

Niels Pinkwart
Humboldt-Universität zu Berlin
Germany

Katrien Verbert
Katholieke Universiteit Leuven
Belgium

LAK 2020 Conference Organization

General Conference Chairs

Christoph Rensing, Technische Universität Darmstadt, GER

Hendrik Drachsler, DIPF Leibniz Institute for Research and Information in Education & University of Frankfurt, GER

Program Chairs

Maren Scheffel, Open Universiteit, NL

Vitomir Kovanović, University of South Australia, AUS

Niels Pinkwart, Humboldt-Universität zu Berlin, GER

Katrien Verbert, KU Leuven, BEL

Practitioner Track Chairs

Jocelyn Manderveld, passion4learning, NL

Bart Rienties, The Open University, UK

Craig Thompson, University of British Columbia, CA

Workshop and Tutorial Chairs

Mar Pérez-Sanagustin, Paul Sabatier University Toulouse, FR

Tobias Ley, Tallinn University, ET

Pedro J. Muñoz-Merino, Universidad Carlos III de Madrid, ES

Doctoral Consortium Chairs

Simon Buckingham-Shum, University of Technology Sydney, AUS

Marcus Specht, TU Delft & LDE Centre for Education and Learning, NL

Stephanie Teasley, University of Michigan, US

Barbara Wasson, University of Bergen, NO

Michael Brown, Iowa State University, US

Posters and Demonstrations Chairs

Jan Schneider, DIPF Leibniz Institute for Research and Information in Education, GER

Daniel Spikol, Malmö University, SE

Yi-Shan Tsai, The University of Edinburgh, UK

Publicity Chairs

Yi-Shan Tsai, The University of Edinburgh, UK

Daniele Di Mitri, Open Universiteit, NL

Sponsorship chair

Jelena Jovanović, University of Belgrade, RS

Proceedings Chairs

Ioana Jivet, Open Universiteit, NL

Atezaz Ahmad, DIPF Leibniz Institute for Research and Information in Education, GER

Website Manager

Tim Steuer, Technische Universität Darmstadt, GER

Organizing Committee

Grace Lynch, Society for Learning Analytics Research, AUS

Nina Seidenberg, studiumdigitale University of Frankfurt, GER

Nicole Hoover, Society for Learning Analytics Research

LAK20 Program Committee

Senior Program Committee

Roger Azevedo, University of Central Florida, USA
Ryan Baker, University of Pennsylvania, USA
Christopher Brooks, University of Michigan, USA
Shane Dawson, University of South Australia, Australia
Hendrik Drachsler, The Open University, The Netherlands
Sidney D'Mello, University of Colorado Boulder, USA
Rebecca Ferguson, The Open University, The United Kingdom
Dragan Gašević, Monash University, Australia
Neil Heffernan, Worcester Polytechnic Institute, USA
Ulrich Hoppe, University Duisburg-Essen, Germany
Srećko Joksimović, University of South Australia, Australia
Jelena Jovanović, University of Belgrade, Serbia
Simon Knight, University of Technology, Sydney, Australia
Manolis Mavrikis, UCL Knowledge Lab, The United Kingdom
Xavier Ochoa, New York University, USA
Hiroaki Ogata, Kyoto University, Japan
Abelardo Pardo, University of South Australia, Australia
Mykola Pechenizkiy, The Eindhoven University of Technology, The Netherlands
Oleksandra Poquet, University of South Australia, Australia
Christoph Rensing, TU Darmstadt, The Netherlands
Bart Rienties, The Open University, The United Kingdom
Carolyn Rosé, Carnegie Mellon University, USA
George Siemens, University of Texas Arlington, USA and University of South Australia, Australia
John Stamper, Carnegie Mellon University, USA
Stephanie Teasley, University of Michigan, USA
Alyssa Wise, Simon Fraser University, Canada
Phil Winne, Simon Fraser University, Canada

Program Committee

Sophie Abel, University of Technology Sydney, Australia
Cecilia Aguerrebere, Duke University, USA
Stephen Aguilar, University of Southern California, USA
Carlos Alario-Hoyos, Universidad Carlos III de Madrid, Spain
Vincent Alevan, Carnegie Mellon University, Pittsburgh, USA
Khulood Alharbi, Durham University, The United Kingdom
Naif Aljohani, University of Southampton, The United Kingdom
Laura Allen, University of New Hampshire, USA
Muhammad Anwar, Virtual University of Pakistan, Pakistan
Kimberly Arnold, University of Wisconsin-Madison, USA
Thushari Atapattu, The University of Adelaide, Australia
Rukmini Avadhanam, University of Minnesota, USA
Aneesha Bakharia, The University of Queensland, Australia
Mohammad Belghis-Zadeh, Athabasca University, Canada
Alan Berg, University of Amsterdam, The Netherlands
Yoav Bergner, New York University, USA
Jason Bernard, University of Saskatchewan, Canada
Sameer Bhatnagar, Polytechnique Montreal, Canada
Daniel Biedermann, German Institute for International Educational Research, Germany
Marion Blumenstein, The University of Auckland, New Zealand
Robert Bodily, Brigham Young University, USA

Anthony F. Botelho, Worcester Polytechnic Institute, USA
 François Bouchet, Sorbonne Université-LIP6, France
 Bert Bredeweg, University of Amsterdam, The Netherlands
 Julien Broisin, Université Toulouse 3 Paul Sabatier - IRIT, France
 Tom Broos, Katholieke Universiteit Leuven, Belgium
 Jill Burstein, Educational Testing Service, USA
 Cristian Cechinel, UFSC-Federal University of Santa Catarina, Brazil
 Rebeca Cerezo, Universidad de Oviedo, Spain
 Mohamed Amine Chatti, University of Duisburg-Essen, Germany
 Bodong Chen, University of Minnesota, USA
 Guanliang Chen, Monash University, Australia
 Weiqin Chen, Oslo and Akershus University College of Applied Sciences, Norway
 Heeryung Choi, University of Michigan, USA
 Irene-Angelica Chounta, University of Tartu, Estonia
 George-Petru Ciordas-Hertel, German Institute for International Educational Research, Germany
 Mihaela Cocea, University of Portsmouth, The United Kingdom
 Cassandra Colvin, The University of Queensland, Australia
 Miguel Ángel Conde, University of León, Spain
 Adam Cooper, Tribal, The United Kingdom
 Linda Corrin, Swinburne University of Technology, Australia
 Alexandra Cristea, Durham University, The United Kingdom
 Yi Cui, Simon Fraser University, Canada
 Mutlu Cukurova, University College London, The United Kingdom
 Crina Damsa, University of Oslo, Norway
 Mihai Dascalu, University Politehnica of Bucharest, Romania
 Daniel Davis, Delft University of Technology, The Netherlands
 Robin De Croon, Katholieke Universiteit Leuven, Belgium
 Maarten De Laat, Open University of the Netherlands, The Netherlands
 Tinne De Laet, Katholieke Universiteit Leuven, Belgium
 Carrie Demmans Epp, University of Alberta, Canada
 Matt Demonbrun, University of Michigan, USA
 Daniele Di Mitri, Open University of the Netherlands, The Netherlands
 Stefan Dietze, GESIS-Leibniz Institute for the Social Sciences, Germany
 Yannis Dimitriadis, University of Valladolid, Spain
 Vania Dimitrova, University of Leeds, The United Kingdom
 Brendan Eagan, University of Wisconsin-Madison, USA
 Michael Eagle, George Mason University, USA
 Vanessa Echeverria, University of Technology Sydney, Australia
 John Erickson, Worcester Polytechnic Institute, USA
 Stephen Fancsali, Carnegie Learning, Inc., USA
 Rafael Ferreira, Federal Rural University of Pernambuco, Brazil
 Ed Fincham, The University of Edinburgh, The United Kingdom
 Barry Fishman, University of Michigan, USA
 James Folkestad, Colorado State University, USA
 Albrecht Fortenbacher, HTW Berlin, Germany
 Joshua Gardner, University of Michigan, USA
 Michail Giannakos, Norwegian University of Science and Technology, Norway
 Andrew Gibson, Queensland University of Technology, Australia
 David Gibson, Curtin University, Australia
 Denis Gillet, Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland
 Art Graesser, University of Memphis, USA
 Wolfgang Greller, Vienna University of Education, Austria
 David Griffiths, University of Bolton, The United Kingdom
 Julio Guerra, University of Pittsburgh, USA
 Gabriel Gutu-Robu, University Politehnica of Bucharest, Romania

Danielle Hagood, University of California, Davis, USA
 Scott Harrison, German Institute for International Educational Research, Germany
 Mohammad Nehal Hasnine, Kyoto University, Japan
 Marek Hatala, Simon Fraser University, Canada
 Claudia Hauff, Germany/University of Technology, The Netherlands
 Tobias Hecking, University of Duisburg-Essen, Germany
 Christothea Herodotou, The Open University, The United Kingdom
 Isabel Hilliger, Pontificia Universidad Católica de Chile, Chile
 Martin Hlosta, The Open University, The United Kingdom
 Tore Hoel, Høgskolen i Oslo og Akershus, Norway
 Caitlin Holman, University of Michigan, USA
 Wayne Holmes, Nesta, The United Kingdom
 Kenneth Holstein, Carnegie Mellon University, USA
 Helen Hong, National Institute of Education, Nanyang Technological University, Singapore
 Geert-Jan Houben, Germany/University of Technology, The Netherlands
 Sarah Howard, University of Wollongong, Australia
 Sharon Hsiao, Arizona State University, USA
 Qian Hu, George Mason University, USA
 Xiao Hu, The University of Hong Kong, Hong Kong
 Dirk Ifenthaler, University of Mannheim, Germany
 Eunice Eunhee Jang, University of Toronto, Canada
 Weijie Jiang, Tsinghua University, China
 Ioana Jivet, Open University of the Netherlands, The Netherlands
 Il-Hyun Jo, Ewha Womans University, Korea
 Terry Judd, The University of Melbourne, Australia
 Sanna Järvelä, University of Oulu, Finland
 Marco Kalz, Heidelberg University of Education, Germany
 Hassan Khosravi, The University of Queensland, Australia
 Gábor Kismihók, Leibniz Information Centre for Science and Technology, Germany
 Kirsty Kitto, University of Technology Sydney, Australia
 René Kizilcec, Cornell University, USA
 Elizabeth Koh, National Institute of Education, Nanyang Technological University, Singapore
 Vitomir Kovanović, The University of South Australia, Australia
 Andy Krumm, Digital Promise, USA
 Ajay Kulkarni, George Mason University, USA
 Tanja Käser, Swiss Data Science Center, USA
 Wing Yin Lam, Technology Enriched Learning Initiative, Hong Kong
 Charles Lang, Columbia University, USA
 Eitel Lauria, Marist College, USA
 Elise Lavoué, iaelyon, Université Jean Moulin Lyon 3, LIRIS, France
 Chi-Un Lei, The University of Hong Kong, Hong Kong
 James Lester, North Carolina State University, USA
 Tobias Ley, Tallinn University, Estonia
 Lisa Lim, University of South Australia, Australia
 Yiwen Lin, University of Michigan, USA
 Danny Y.T. Liu, The University of Sydney, Australia
 Lori Lockyer, University of Technology Sydney, Australia
 Jason Lodge, The University of Queensland, Australia
 Steven Lonn, University of Michigan, USA
 Rose Luckin, The London Knowledge Lab, The United Kingdom
 Vanda Luengo, Sorbonne Université-LIP6, France
 Leah Macfadyen, The University of British Columbia, Canada
 Armando Maciel-Toda, TODO
 Katerina Mangaroska, Norwegian University of Science and Technology, Norway
 Bannert Maria, Technical University of Munich, Germany

Roberto Martínez-Maldonado, Monash University, Australia
 Alejandra Martínez-Monés, Universidad de Valladolid, Spain
 Claudia Mazziotti, Technical University of Munich, Germany
 Jenny McDonald, The University of Auckland, New Zealand
 Agathe Merceron, Beuth University of Applied Sciences Berlin, Germany
 Nikola Milikić, University of Belgrade, Serbia
 Martijn Millecamp, Katholieke Universiteit Leuven, Belgium
 Sandra Milligan, The University of Melbourne, Australia
 Caitlin Mills, University of New Hampshire, Canada
 Negin Mirriahi, University of South Australia, Australia
 Inge Molenaar, Radboud University, The Netherlands
 Pedro Manuel Moreno-Marcos, Universidad Carlos III de Madrid, Spain
 Benjamin Motz, Indiana University Bloomington, USA
 Kousuke Mouri, Tokyo University of Agriculture and Technology, Japan
 Pablo Munguia, RMIT University, Australia
 Hanni Muukkonen, University of Oulu, Finland
 Pedro J. Muñoz-Merino, Universidad Carlos III de Madrid, Spain
 Tzi Dong Jeremy Ng, The University of Hong Kong, Hong Kong
 Le Quan Nguyen, University of Michigan, USA
 Katja Niemann, XING SE, Germany
 Jaclyn Ocumpaugh, University of Pennsylvania, USA
 Zacharoula Papamitsiou, Norwegian University of Science and Technology, Norway
 Luc Paquette, University of Illinois at Urbana-Champaign, USA
 Abelardo Pardo, University of South Australia, Australia
 Zachary Pardos, University of California Berkeley, USA
 Yeonjeong Park, Honam University, Korea
 Melanie Peffer, University of Colorado Boulder, USA
 Mar Pérez-Sanagustín, Pontificia Universidad Católica de Chile, Chile
 Carlo Perrotta, Monash University, Australia
 Héctor Javier Pijeira-Díaz, University of Oulu, Finland
 Sasha Poquet, University of South Australia, Australia
 Luis P. Prieto, Tallinn University, Estonia
 Tiago Primo, Federal University of Pelotas, Brazil
 Paul Prinsloo, University of South Africa, South Africa
 Joshua Quick, Indiana University Bloomington, USA
 Mladen Raković, The University of North Carolina at Chapel Hill, USA
 Vinicius Ramos, Universidade Federal de Santa Catarina, Brazil
 Huzefa Rangwala, George Mason University, USA
 Justin Reich, Massachusetts Institute of Technology, USA
 Peter Reimann, University of Sydney, Australia
 Zhiyun Ren, The Ohio State University, USA
 Ulla Ringtved, University College of Northern Denmark, Denmark
 Steven Ritter, Carnegie Learning, Inc., USA
 Covadonga Rodrigo, UNED-Universidad Nacional de Educacion a Distancia, Spain
 Fernando Rodriguez, University of California, Irvine, USA
 María Jesús Rodríguez-Triana, Tallinn University, Estonia
 Ido Roll, Technion-Israel Institute of Technology, Israel
 Tamra Ross, Athabasca University, Canada
 José A. Ruipérez Valiente, Massachusetts Institute of Technology, USA
 Demetrios Sampson, Curtin University, Australia
 Maria Ofelia San Pedro, ACT, Inc., USA
 Agnes Sandor, Naver Labs Europe, France
 Olga C. Santos, aDeNu Research Group (UNED), Spain
 Mohammed Saqr, University of Eastern Finland, Finland
 Marcel Schmitz, Zuyd Hogeschool, The Netherlands

Jan Schneider, DIPF Leibniz Institute for Research and Information in Education, Germany
 Niall Sclater, Sclater Digital, The United Kingdom
 Paul Seitlinger, Tallinn University, Estonia
 Mike Sharkey, Data & Graphs, USA
 Kshitij Sharma, Norwegian University of Science and Technology, Norway
 Bruce Sherin, Northwestern University, USA
 Antonette Shibani, University of Technology Sydney, Australia
 Atsushi Shimada, Kyushu University, Japan
 Erica Snow, Imbellus, USA
 Sergey Sosnovsky, Utrecht University, The Netherlands
 Marcus Specht, Delft University of Technology, LDE-CEL, The Netherlands
 Daniel Spikol, Malmö University, Sweden
 Tamara Sumner, University of Colorado Boulder, USA
 Zachari Swiecki, University of Wisconsin-Madison, USA
 Kairit Tammets, Tallinna Ülikool, Estonia
 Yuta Taniguchi, Kyushu University, Japan
 Michelle Taub, University of Central Florida, USA
 Stefaan Ternier, Open University of the Netherlands, The Netherlands
 Craig Thompson, The University of British Columbia, Canada
 Stefan Trausan-Matu, University Politehnica of Bucharest, Romania
 Yi-Shan Tsai, The University of Edinburgh, The United Kingdom
 Anouschka van Leeuwen, Utrecht University, The Netherlands
 Manuel Valle Torre, Germany/ft University of Technology, The Netherlands
 Lorenzo Vigentini, The University of New South Wales, Australia
 Jovita Vytasek, Simone Fraser University, Canada
 Elle Yuan Wang, Arizona State University, USA
 Denise Whitelock, The Open University, The United Kingdom
 Alexander Whitelock-Wainwright, Monash University, Australia
 Annika Wolff, Lappeenranta University of Technology, Finland
 Sebastian Wollny, DIPF German Institute for International Educational Research, Germany
 Marcelo Worsley, Northwestern University, USA
 Masanori Yamada, Kyushu University, Japan
 Jialin Yu, Durham University, The United Kingdom
 Renzhe Yu, University of California, USA
 Zdenek Zdrahal, The Open University, The United Kingdom
 Mengxiao Zhu, Educational Testing Service, USA
 Amal Zouaq, University of Ottawa, Canada

Additional Reviewers

Abel, Sophie
 Alharbi, Khulood
 Anwar, Muhammad
 Avadhanam, Rukmini
 Belghis-Zadeh, Mohammad
 Bernard, Jason
 Bhatnagar, Sameer
 Biedermann, Daniel
 Ciordas-Hertel, George-Petru
 Eagan, Brendan
 Erickson, John

Gutu-Robu, Gabriel
 Hasnine, Mohammad Nehal
 Hilliger, Isabel
 Hoel, Tore
 Hu, Qian
 Jiang, Weijie
 Jo, Il-Hyun
 Kulkarni, Ajay
 Lam, Wing Yin
 Lin, Yiwen
 Maciel-Toda, Armando
 Moreno-Marcos, Pedro Manuel

Ng, Tzi Dong Jeremy
 Primo, Tiago
 Quick, Joshua
 Ramos, Vinicius
 Ren, Zhiyun
 Ross, Tamra
 Schmitz, Marcel
 Seitlinger, Paul
 Valle Torre, Manuel
 Wollny, Sebastian
 Yu, Jialin

Table of Contents

Practitioner reports

Implementing an attrition model at scale <i>Amelia Brennan, Andras Nemes, Jeevan Pokhrel, Cameron Doyle, Vishesh Jain and David McLay</i>	1–4
Effects of In-class and Out-of-class Learning Behaviors on Learning Performance and Self-regulated Learning Awareness <i>Li Chen, Yoshiko Goda, Atsushi Shimada and Masanori Yamada</i>	5–8
Learning Analytics Applied in Commercial Aviation: Our Use Cases, Obstacles, and Help Needed From Academia <i>Liz Gehr and Laurie Dunagan</i>	9–12
For Evidence-Based Class Design with Learning Analytics: A Proposal of Preliminary Practice Flow Model in High School <i>Satomi Hamada, Yufan Xu, Xuewang Geng, Li Chen, Hiroaki Ogata, Atsushi Shimada and Masanori Yamada</i>	13–16
Building a Unified Data Platform to Support Metacognition and Self Regulated Learning <i>John Johnston and Etienne Pelaprat</i>	17–20
Using the Behaviour Change Wheel for Learning Analytics adoption <i>Hazel Jones</i>	21–24
Building a data warehouse for multimodal learning analytics research projects <i>Vlatko Lukarov, Matthias Ehlenz and Ulrik Schroeder</i>	25–28
Engaging Students as Co-Designers of Learning Analytics <i>Juan Pablo Sarmiento, Fabio Campos and Alyssa Wise</i>	29–32
Considerations for amending a whole-institution early-alert system <i>Rebecca Siddle and Ed Foster</i>	33–36
Why clicks are not enough: designing a Learning Analytics service for the Estonian Open Educational Resources ecosystem <i>Kairit Tammets, Tobias Ley, Mart Laanpere and Manisha Khulbe</i>	37–40

Research Posters

Open Learning Analytics Indicator Repository <i>Atezaz Ahmad, Jan Schneider and Hendrik Drachsler</i>	41–43
Examining the Relationship Between Temporality and Social Positions in Social Annotation <i>Rukmini Manasa Avadhanam and Bodong Chen</i>	44–46
Barriers and Hurdles to the Publication of Learning Analytics Data <i>Katarzyna Biernacka and Niels Pinkwart</i>	47–49
Physiology-Aware Learning Analytics Using Pedagogical Agents <i>Melanie Bleck, Nguyen-Thinh Le and Niels Pinkwart</i>	50–52
Exploring Writing Achievement and Genre in Postsecondary Writing <i>Jill Burstein, Daniel McCaffrey, Norbert Elliot and Beata Beigman Klebanov</i>	53–55
Assessing Risk in Learning Analytics Projects <i>Henrique Chevreux, Valeria Henríquez, Eliana Scheihing, Pedro Muñoz-Merino, Tinne de Laet, Mar Pérez-Sanagustín, Isabel Hilliger, Jorge Maldonado-Mahauad, Paola Pesantez and Margarita Ortíz</i>	56–58

EduBERT: Deep Language Models for Learning Analytics <i>Benjamin Clavié and Kobi Gal</i>	59–61
Using Eye Tracking Data to Analyze the Effects of Learning Hints in Source Code Comprehension <i>Fabian Deitelhoff, Andreas Harrer, Benedikt Schröder, Ulrich Hoppe and Andrea Kienle</i>	62–64
Semantically Adjusting Word Frequency for Estimating Word Difficulty from Unbalanced Corpora <i>Yo Ehara</i>	65–67
Adaptive Learning Guidance System (ALGS) <i>Ghada El-Hadad, Doaa Shawky and Ashraf Badawi</i>	68–70
Teachers’ Sense-Making of Learning Analytics: what is missing? <i>Diana Forero Tobon, Canan Blake, Yvonne Vezzoli, Mina Vasalou and Manolis Mavrikis</i>	71–73
A proactive perspective on the future of Learning Analytics: A systematic literature review <i>Stephanie Gaaw and Cathleen M. Stuetzer</i>	74–76
Using Writing Logs to Help Validate Assessments for Learning <i>Hongwen Guo, Mo Zhang, Paul Deane and Randy Bennett</i>	77–79
Why Predictions of At-Risk Students Are Not 100% Accurate? Showing Patterns in False Positive and False Negative Predictions <i>Martin Hlosta, Zdenek Zdrahal, Vaclav Bayer and Christothea Herodotou</i>	80–82
Describing Teachers’ Self-Regulation of Information Seeking: Preliminary Results from Physiological Arousals and Log Files <i>Lingyun Huang and Susanne Lajoie</i>	83–85
Temporal analytics of log data derived from students’ manipulating mathematical objects <i>Masataka Kaneko, Takahiro Nakahara and Takeo Noda</i>	86–88
Using network analysis to investigate the relation between knowledge organization and transfer <i>Marcus Kubsch</i>	89–91
Accurate and Interpretable Sensor-free Affect Detectors via Monotonic Neural Networks <i>Andrew Lan, Anthony Botelho, Shamyia Karumbaiah, Ryan Baker and Neil Heffernan</i>	92–94
Online Learning Diagnosis and Feedback System Using Academic Emotion Data <i>Jihyang Lee and Hyo-Jeong So</i>	95–97
Triangulating Multimodal Data to Measure Self-Regulated Learning <i>Kia Puay Lim, Joep van der Graaf, Yizhou Fan, Katharina Engelmann, Maria Bannert, Inge Molenaar, Dragan Gasevic and Johanna Moore</i>	98–100
How do the Game Level Plateaus Inform the Learning Design? <i>Anna Lizarov, Charles Lang and Kara Carpenter</i>	101–103
Development of a Learning Dashboard Prototype Supporting Meta-cognition for Students <i>Min Lu, Li Chen, Yoshiko Goda, Atsushi Shimada and Masanori Yamada</i>	104–106
From Text Mining to Evidence Team Learning in Cybersecurity Exercises <i>Kaie Maennel, Joonsoo Kim and Stefan Sütterlin</i>	107–109
LA Platform in Junior High School: Trends of Usage and Student Performance <i>Rwitajit Majumdar, Hiroyuki Kuromiya, Kiriko Komura, Brendan Flanagan and Hiroaki Ogata</i>	110–112
Analysis of Students’ Self-Regulatory Strategies in MOOCs <i>Jorge Maldonado-Mahauad and Mar Perez-Sanagustin</i>	113–115

Implementing learning analytics in schools: Towards a theory-based and data-driven framework for analytics implementation <i>Claudia Mazziotti, Vitomir Kovanovic, Shane Dawson and George Siemens</i>	116–118
Analytics of multimodal learning logs for page difficulty estimation <i>Tsubasa Minematsu, Atsushi Shimada and Rin-Ichiro Taniguchi</i>	119–121
Reduction of Supervised Data for Chat Analysis in Collaborative Learning by Using Transfer Learning Methods <i>Shun Moriya, Chihiro Shibata, Kimihiko Ando and Taketoshi Inaba</i>	122–124
Precision-Based Predictive Analytics <i>Patsy Moskal, Thomas Cavanagh, Morgan Wang and Jianbin Zhu</i>	125–127
Toward Predicting Learners' Efficiency for Adaptive e-Learning <i>Yuta Nakashima, Hirokazu Kobori, Ryota Takaoka, Noriko Takemura, Tsukasa Kimura, Hajime Nagahara, Masayuki Numao and Kazumitsu Shinohara</i>	128–130
Simultaneously Learning Competencies and Item Difficulties <i>Kai Neubauer and Ulf Brefeld</i>	131–133
A Qualitative Analysis of One University's Ethical Fears and Practical Desires for Learning Analytics <i>Melanie Peffer and Jessie Sutton</i>	134–136
Automated Essay Scoring in Foreign Language Students Based on Deep Contextualised Word Representations <i>Bojana Ranković, Sarah Smirnow, Martin Jaggi and Martin Tomasik</i>	137–139
Scaffolding Teachers Sense Making for Classroom Equity using Visual Analytics <i>Ali Raza, William Penuel, Jennifer Jacobs and Tamara Sumner</i>	140–142
A Note on Automatic Grading of Short Answers and Providing Feedback <i>Neslihan Suzen, Alexander N Gorban, Jeremy Levesley and Evgeny M Mirkes</i>	143–145
Chatbots – An Opportunity for Individual Assistance in Education <i>Sebastian Wollny, Jan Schneider, Marc Rittberger and Hendrik Drachler</i>	146–148
De-identification is not enough to guarantee student privacy: De-anonymizing personal information from basic logs <i>Elad Yacobson, Orly Fuhrman, Sara HersHKovitz and Giora Alexandron</i>	149–151
What Do You Get from Replies? Causal Estimates of Peer Effects in Online Discussion Forums <i>Renzhe Yu and Di Xu</i>	152–154
Predicting College Success: What Data Are Useful and for Whom? <i>Renzhe Yu, QiuJie Li, Christian Fischer, Di Xu and Shayan Doroudi</i>	155–157
Profiling students' mathematics motivation in 9th-grade: How does student motivation change over one academic semester? <i>Xiaoxue Zhang, James Middleton, Elizabeth Farley-Ripple, Zachary Collier and Amanda Jansen</i>	158–160
Personality-Aware Educational Recommendations: Subjective vs Inferred Personality Traits <i>Yong Zheng and Archana Subramaniyan</i>	161–163
Practitioner Posters	
Learning Analytics for Inclusive STEM Student Success <i>G. Alex Ambrose, Xiaojing Duan, Kevin Abbott, Victoria Woodard and Kelley Young</i>	164–165

Designing A Collaborative Problem-Solving Activity to Prepare Students for Flipped Classroom <i>Yuqian Chai, Xinyu Qi, Ling Li, Mansurbek Kushnazarov, Cheuk-Wang Yau, Yifei Dong and Chi-Un Lei</i>	166–167
Towards Better Grading: Promoting Grading Fairness with Assessment Decision Tree Visualization <i>Yuqian Chai, Xinyu Qi, Ling Li, Mansurbek Kushnazarov, Cheuk-Wang Yau, Yifei Dong and Chi-Un Lei</i>	168–169
Fertile breeding ground for learning analytics at scale: the KU Leuven approach <i>Tinne De Laet, Tom Broos, Inge Wullaert, Anneleen Cosemans and Katleen Craenen</i>	170–171
Towards Instructor-based Predictive Learning Analytics <i>Jesse Eickholt and Chris Phillips</i>	172–173
Getting staff ready for learning analytics: preliminary findings from a 3-year project <i>Ed Foster, Rebecca Siddle and Pieterjan Bonne</i>	174–175
Student-centered Development of Learning Analytics at an Higher Education Institution <i>Jiri Lallimo and Amanda Sjöblom</i>	176–177
Towards a modular and flexible Learning Analytics framework <i>Yves Noël, Roland Mergoïl, Vanda Luengo and François Bouchet</i>	178–179
Development of a Data-Oriented e-Learning Platform Based on the Community of Inquiry Framework <i>Xinyu Qi, Yuqian Chai, Ling Li, Mansurbek Kushnazarov, Cheuk-Wang Yau, Yifei Dong and Chi-Un Lei</i>	180–181
Supporting online learners' awareness through weekly study reports for self-regulation of learning <i>Min Qu, Héctor J. Pijera-Díaz and Yutian Ma</i>	182–183
The use of an AI algorithm to verify exam questions on prescriptions within e-learning programme P-scribe <i>Floor van Rosse and Adriaan van Doorn</i>	184–185

Practitioner Demonstrations

CoTrack: A Tool for Tracking Collaboration Across Physical and Digital Spaces in Collocated Blended Settings <i>Pankaj Chejara, Adolfo Ruiz-Calleja, Luis P. Prieto, María Jesús Rodríguez-Triana and Shashi Kant Shankar</i>	186–186
Creating an Interface Supported by Learning Analytics to Assist Learners with Navigating Individualized Learning Pathways <i>Matt Crosslin</i>	187–187
Student advising learning dashboards: the story of LISSA and LALA <i>Tinne De Laet, Tom Broos, Martijn Millecamp, Katrien Verbert, Julio Guerra, Margarita Ortiz and Miguel Angel Zuniga-Pietro</i>	188–188
PerformanceVis: Homework & Exam Analytics Dashboard for Inclusive Student Success <i>Xiaojing Duan, Alex Ambrose, Chaoli Wang, Kevin Abbott, Victoria Woodard and Catlin Schalk</i>	189–189
MINDSTEPS: An Adaptive Computer-Based Tool for Formative Student Assessment <i>Nina König, Martin J. Tomasik, Stéphanie Berger, Lukas Giesinger, Laura A. Helbling and Urs Moser</i>	190–190
FLOWer: Feedback Loop for Co-located Group-work Support <i>Yoon Lee, Haoyu Chen, Esther Tan, Sambit Praharaj and Marcus Specht</i>	191–191
The Fellowship of the Learning Activity: playing, cooperating, creating awareness and designing learning activities <i>Marcel Schmitz and Maren Scheffel</i>	192–192
BOA: Learner Analytics as an Advising Resource <i>Steven Williams</i>	193–193

Doctoral Consortium

A novel feedback system for refinement and improvement of lecture-based pedagogy <i>Pankaj Chavan</i>	194–200
Modelling Student Participation Using Discussion Forum Data <i>Elaine Farrow</i>	201–206
Guidance in Multimodal Learning Analytics for Collaborative Classrooms <i>Gloria Milena Fernandez-Nieto</i>	207–212
Learning Analytics and Capability Approach in Education: Analysing Student Agency in Higher Education <i>Ville Heilala</i>	213–218
Discovering Teachers' Regulatory Learning Processes in the Context of Technology Integration Using Educational Data Mining Approaches <i>Lingyun Huang and Susanne Lajoie</i>	219–224
The Importance of Feedback Actionability in Learning Environments <i>Hamideh Iraj</i>	225–229
Unpacking the Bi-directional relationship between learning analytics and learning design in blended learning environments <i>Rogers Kaliisa</i>	230–235
Detecting patterns of self-regulated learning in a virtual classroom environment <i>Madiha Khan</i>	236–241
The Institutional Logics of Learning Analytics in Higher Education: Implications for Organizational and Individual Outcomes <i>Carrie Klein</i>	242–247
Understanding Metacognition in Online Learning: To What Extent Learners Trajectories Converge with Self-Reported Behaviors <i>Fatemeh Salehian Kia, Marek Hatala and Ryan S. Baker</i>	248–253
AI Driven Support System (ADSS) for Teachers to Increase Underrepresented Minorities Academic Success <i>Taylor Stevens</i>	254–259

Workshops

The 6th LAKathon: Accelerating Development by Learning from the Past <i>Daniele Di Mitri, Gábor Kismihók, Alan Berg, Kirsty Kitto, Stefan Mol, Jan Schneider and Jose Ruipérez-Valiente</i>	260–263
Linking Individual Contribution Assessment in Collaborative Learning with Learning Analytics: A Multimodal Approach <i>Song Lai, Hao Tian, Jiaqi Liu and Fati Wu</i>	264–265
Assessing Learning in Real Time – Learning Analytics strategy for classrooms <i>Vijayan N</i>	266–267
Learning Analytics for Classroom Activities <i>Jan Schneider, Daniele Di Mitri and George-Petru Ciordas-Hertel</i>	268–268
Applicability of Automatic Short Answer Grading Systems in Assessment Scenarios <i>Anna Filighera and Tim Steuer</i>	269–270

Benchmarking Ethical LA algorithms <i>Alan Mark Berg, Shereif Eid, Gábor Kismihók and Stefan T. Mol</i>	271–272
Supporting Learning Design Activities Outside the Classroom with Learning Analytics <i>Gerti Pishtari, María Jesús Rodríguez-Triana</i>	273–274
Data-driven Assessment of Spatial Reasoning in a Geometry Game through Learning Analytics <i>José A. Ruipérez Valiente and Yoon Jeon Kim</i>	275–275
Smartwatches in Education <i>Daniele Di Mitri, Khaleel Asyraaf Mat Sanusi and Jan Schneider</i>	276–277
Adoption, Adaptation and Pilots of Learning Analytics for Latin American Higher Education Institutions <i>Pedro J. Muñoz-Merino, Carlos Delgado Kloos, Yi-Shan Tsai, Dragan Gasevic, Katrien Verbert, Mar Perez-Sanagustin, Isabel Hilliger, Miguel Ángel Zúñiga Prieto, Margarita Ortiz and Eliana Scheihing</i>	278–281
2nd Workshop on Learning Analytic and Services to Support Personalized Learning and Assessment at Scale <i>Alina Von Davier, Michael Yudelson, John Stamper, Steven Ritter and Peter Brusilovsky</i>	282–284
Teacher-sourcing semantic information in a Physics blended learning environment <i>Giora Alexandron, Elad Yacobson, Asaf Bar Yosef and Eliran Hen</i>	285–292
Understanding Reflective Writing Criteria in Computer Science Education from CS Educators in Higher Education <i>Huda Alrashidi, Thomas Daniel Ullmann, Samiah Ghounaim and Mike Joy</i>	293–301
Facilitating Adaptive Assessment Delivery at Scale: Echo-Adapt Software-As-A-Service <i>Michelle Barrett and Bingnan Jiang</i>	302–311
Providing Directed Feedback Through QUICK-Comments <i>Anthony F. Botelho, John A. Erickson, Aaron G. Alphonsus and Neil T. Heffernan</i>	312–321
Opportunities for Human-AI Collaborative Tools to Advance Development of Motivation Analytics <i>Steven C. Dang and Kenneth R. Koedinger</i>	322–329
Sphinx: An Automated Generation System for English Reading Comprehension Assessment <i>Saad Khan, Yuchi Huang, Scott Pu, Vladimir Tarasov, Alejandro Andrade, Richard Meisner, Dave Edwards and Alina von Davier</i>	330–340
Adaptive Learning Meets Crowdsourcing: Towards Development of Cost-Effective Adaptive Educational Systems <i>Hassan Khosravi</i>	341–347
The Modeling of Digital Learning Networks <i>Oleksandra Poquet, Tobias Hecking and Bodong Chen</i>	348–351
Exploration of Peer Effects through Digital Forum Interactions <i>Oleksandra Poquet</i>	352–354
The Current State of Knowledge-Building Analytics and Possible Future Directions <i>Jun Oshima and Ritsuko Oshima</i>	355–357
Modelling students' social network structure from spatial-temporal network data <i>Quan Nguyen, Warren Li and Christopher Brooks</i>	358–360
Tree Structure of Collective Attention Network: Revisiting the Problem of Dropout <i>Jingjing Zhang and Ming Gao</i>	361–363

The Flow of Books in the Era of Social Media: A case study of a reading group <i>Lei Zhang and Xiangdong Chen</i>	364–366
The ACT Master(y) model for Measurement, Learning and Navigation <i>Gunter Maris, Steve Polyak, Michael Yudelson</i>	367–370
Evidence Based Decision Making in the Classroom <i>Geraldine Gray, Pauline Rooney, James Doody, Phelim Murnion, Kevin O’Rourke, Lee O’Farrell, Charles Lang</i>	371–374
DesignLAK20: Developing quality standards for analytic measures of learning for learning design <i>Sandra Milligan, Linda Corrin, Nancy Law and Ulla Ringtved</i>	375–378
Building Capacity Through the Learning Analytics Learning Network <i>Justin Dellinger, Florence Gabriel, Ryan Baker, Shane Dawson and George Siemens</i>	379–381
Learning Analytics Principles of Use: Making Ethics Actionable <i>Kimberly Arnold, Marcia Ham, Robin Pappas and George Rehrey</i>	382–385
3rd Personalising feedback at scale Workshop: Teacher-driven action and dialogue <i>Lorenzo Vigentini, Sarah Howard, Danny Y.T. Liu and Lisa Lim</i>	386–389
Analytics-based Personalised Feedback and Learning Outcomes in ESL at School Level <i>Deepti Yadav and N. C. Ojha</i>	390–394
‘It’s a good tool, just hard to use’: Navigating staff perceptions during the institution-wide implementation of OnTask <i>Sheridan Gentili and Amanda Richardson</i>	395–402
Social Network-based Awareness Tools in Collaborative Learning <i>Lei Zhang, Xiangdong Chen</i>	403–408
Toward the Design of a Learning Analytics for Learning Design Dashboard in Location-based Learning <i>Gerti Pishtari, María Jesús Rodríguez-Triana</i>	409–412
Learning Analytics Support for Dialogic Peer Feedback <i>Erkan Er, Yannis Dimitriadis and Dragan Gašević</i>	413–419
What matters?: The role of feedback in developing preservice teachers’ digital competency for future teaching <i>Sarah K. Howard, Jo Tondeur and Jack Yang</i>	420–424
Integrating Multi-channel Learning Data to Model Complex Learning Processes <i>Roger Azevedo, George Siemens and Shane Dawson</i>	425–428
Contextualizing multimodal learning analytics to theoretical frameworks and learning environments <i>Elizabeth B. Cloude and Roger Azevedo</i>	429–432
Measuring Micro-Level Self-Regulated Learning Processes with Enhanced Log Data and Eye Tracking Data <i>Y. Fan, K.P. Lim, J. van der Graaf, J. Kilgour, K. Engelmann, M. Bannert, I. Molenaar, J. Moore and D. Gasevic</i>	433–436
Identifying Trigger Regulation Events in Collaborative Learning <i>Sanna Järvelä, Jonna Malmberg, Eetu Haataja and Muhterem Dindar</i>	437–441
Finding Semantic Structure of Content from Gaze Data <i>Hiroaki Kawashima</i>	442–445
Mobile Multimodal Learning Analytics Method to Foster Student Self-Regulated Learning <i>Mohammad Khalil and Olga Viberg</i>	446–449

Fifth Grade Students' Problem-Solving Strategies and "Aha! Moments" in Authentic Informal STEM Environment <i>Yaoran Li, Vitaliy Popov, Perla Myers, Joi Spencer, Odesma Dalrymple and Scott Lundergan</i>	450–454
Challenges in Multichannel Data Discovery and Integration for Monitoring Performance in Self-Regulated Learning <i>Shashi Kant Shankar</i>	455–458
Collecting and Integrating Multimodal Data from a Programming Exercise Environment <i>Yuta Taniguchi and Atsushi Shimada</i>	459–462
EdRecSys Workshop@LAK2020 <i>Martin Hlostá, Christopher Krauss, Katrien Verbert, Geoffray Bonnin, Martijn Millecamp, Vaclav Bayer</i>	463–466
Find the Optimizing Time intervals in Programming Learning for College Students <i>Fangjing Ning, Baoping Li, Penghe Chen, Qiuyu Chen and Xianru Zhang</i>	467–469
Mini Survival Kit: Prediction based recommender to help students escape their critical situation in online courses <i>Martin Hlostá, Vaclav Bayer and Zdenek Zdrahal</i>	470–472
Culturally inclusive learning analytics (#CILA) <i>Ioana Jivet, Tom Broos, Hendrik Drachsler and Maren Scheffel</i>	473–476
XLA: Explainable Learning Analytics <i>Katrien Verbert, Tinne De Laet, Martijn Millecamp, Tom Broos, Mohamed Amine Chatti and Arham Muslim</i>	477–479
A Review of Explanatory Visualizations in Recommender Systems <i>Mouadh Guesmi, Mohamed Amine Chatti, Arham Muslim</i>	480–491
Exploring the Design Space for Explainable Course Recommendation Systems in University Environments <i>Boxuan Ma, Min Lu, Yuta Taniguchi, Shin'ichi Konomi</i>	492–499
Explainable Learning Analytics: challenges and opportunities <i>Tinne De Laet, Martijn Millecamp, Tom Broos, Robin De Croon, Katrien Verbert, Raphael Duorado</i>	500–510
Teacher-Centered Dashboards Design Process <i>Mohamed Ez-zaouia</i>	511–528
CrossMMLA in practice: Collecting, annotating and analyzing multimodal data across spaces <i>Michail Giannakos, Daniel Spikol, Inge Molenaar, Daniele Di Mitri, Kshitij Sharma, Xavier Ochoa and Rawad Hammad</i>	529–532
Context-aware Multimodal Learning Analytics Taxonomy <i>Maka Eradze, María Jesús Rodríguez Triana and Mart Laanpere</i>	533–538
Towards Teacher Orchestration Load-aware Teacher-facing Dashboards <i>Ishari Amarasinghe, Milica Vujovic and Davinia Hernández-Leo</i>	539–542
MMLA Approach to Track Collaborative Behavior in Face-to-Face Blended Settings <i>Pankaj Chejara, Reet Kasepalu, Shashi Kant Shankar, Luis P. Prieto, María Jesús Rodríguez- Triana and Adolfo Ruiz-Calleja</i>	543–548
Physiology-Aware Learning Analytics Using Pedagogical Agents <i>Melanie Bleck, Nguyen-Thinh Le and Niels Pinkwart</i>	549–554
Designing and implementing multimodal data collection in classroom to capture metacognition in collaborative learning <i>Jonna Malmberg, Sanna Järvelä, Hanna Järvenoja, Eetu Haataja, Héctor J. Pijera-Díaz, Ahsen Cini</i>	555–561

Multimodal Temporal Network Analysis to Improve Learner Support and Teaching <i>Mohammed Saqr, Olga Viberg, Jalal Nouri, Solomon Oyelere</i>	562–565
Using Multimodal Learning Analytics to Explore how Children Experience Educational Motion-Based Touchless Games <i>Serena Lee-Cultura, Kshitij Sharma, Michail Giannakos</i>	566–571
The challenge of interaction assignment on large digital tabletop displays for learning analytics <i>Matthias Ehlenz, Vlatko Lukarov, Ulrik Schroeder</i>	572–577
Facilitating Self-Regulated Learning with Personalized Scaffolds on Student’s own Regulation Activities <i>Joep van der Graaf, Inge Molenaar, Lyn Lim, Yizhou Fan, Katharina Engelmann, Dragan Gašević and Maria Bannert</i>	578–580
Endowing Head-Mounted Displays with Physiological Sensing for Augmenting Human Learning and Cognition <i>Evangelos Niforatos, Athanasios Vourvopoulos and Michail Giannakos</i>	581–584
Towards Collaboration Literacy Development through Multimodal Learning Analytics <i>Marcelo Worsley and Xavier Ochoa</i>	585–595
SPARK: A Learning Analytics Leadership Framework <i>Shane Dawson, Abelardo Pardo and George Siemens</i>	596–599
Quantitative ethnography as a framework for network analysis – a discussion of the foundations for network approaches to leaning analysis <i>Morten Misfeldt, Daniel Spikol, Jesper Bruun, Mohammed Saqr, Rogers Kaliisa, Andrew Ruis, Brendan Eagan and Kamila Misiejuk</i>	600–603
The 2nd Workshop on Predicting Performance Based on the Analysis of Reading Behavior <i>Brendan Flanagan, Rwitajit Majumdar, Atsushi Shimada and Hiroaki Ogata</i>	604–607
Improving Learning Analytics and Student Performance through Connected Lifelong Learning on the Blockchain <i>Ocheja Patrick, Flanagan Brendan, Lecailliez Louis and Ogata Hiroaki</i>	608–616
Design of an Interactive Dashboard for Teacher Orchestration in Collaborative Science Inquiry <i>Jiaxin Cao and Yanjie Song</i>	617–622
Understanding Jump Back Behaviors in E-book System <i>Boxuan Ma, Jiadong Chen, Chenhao Li, Likun Liu, Min Lu, Yuta Taniguchi, Shin’ichi Konomi</i>	623–631
Social Knowledge Mapping Tool for Interactive Visualization of Learners' Knowledge <i>Akira Onoue, Masanori Yamada, Atsushi Shimada, Tsubasa Minematsu, Rin-ichiro Taniguchi</i>	632–637
Can the Area marked in eBook Readers Specify Learning Performance? <i>Xu Yufan, Geng Xuewang, Chen Li, Hamada Satomi, Taniguchi Yuta, Ogata Hiroaki, Shimada Atsushi, Masanori Yamada</i>	638–648
Recommendation of Personalized Learning Materials based on Learning History and Campus Life Sensing <i>Keita Nakayama, Atsushi Shimada, Tsubasa Minematsu, Masanori Yamada, Rin-ichiro Taniguchi</i>	649–654
OpenLA: An Open-Source Library for e-Book Log Analytics <i>Ryusuke Murata, Tsubasa Minematsu, Atsushi Shimada</i>	655–660
Learning Engagement - Clustering Analysis based on Student Interaction with Digital Textbooks <i>Abu Abu Eyo, Owoeye Oluwaseyi, Ocheja Patrick, Flanagan Brendan, and Ogata Hiroaki</i>	661–668

Learner's Performance Prediction based on Histogram of Actions during lecture <i>Takayoshi Yamashita, Tsubasa Hirakawa, Hironobu Fujiyoshi</i>	669–674
Score Prediction Based on Page Feature Clustering <i>Ryusuke Murata, Tsubasa Minematsu, Atsushi Shimada</i>	675–677
Performance prediction using dimension-reduced activity data <i>Taisei Aoki, Yuya Kida, Maiya Hori</i>	678–679
Predicting Student Exam Scores Based on Click-stream Level Data of Their Usage of an E-Book System <i>Makhlouf Jihed and Tsunenori Mine</i>	680–689
A Picture-Book Recommender System for Extensive Reading on an E-Book System <i>Chifumi Nishioka, Sanae Fujita, Takashi Hattori, Tessei Kobayashi, Futoshi Naya, Hiroaki Ogata</i>	690–692
What Activity Contributes to Academic Performance? <i>Tetsuya Shiino, Tsubasa Minematsu, Atsushi Shimada, Rin-ichiro Taniguchi</i>	693–695
Automatic Retrieval of Learning Contents Related to Quizzes for Supporting Students' Enhanced Reviews <i>Takashi Ishikawa, Tsubasa Minematsu, Atsushi Shimada, Rin-ichiro Taniguchi</i>	696–698
Generating individual advice corresponding to the learning level by analyzing learning behaviors <i>Taisei Aoki, Maiya Hori, Atsushi Shimada</i>	699–701
Evaluating the Accuracy of Real-time Learning Analytics in Student Activities <i>Takuro Owatari, Atsushi Shimada, Tsubasa Minematsu, Rin-ichiro Taniguchi</i>	702–704
LAK Theory 2020: Workshop on Theory and Learning Analytics <i>Kathryn Bartimote, Sarah Howard and Dragan Gasevic</i>	705–708
Let's Talk LA: Discussing Challenges for Institutional Adoption of Learning Analytics <i>Isabel Hilliger, Yi-Shan Tsai, Pedro J. Muñoz Merino and Mar Pérez-Sanagustín</i>	709–712
Bringing together writing tool design, writing analytics and writing pedagogy <i>Christian Rapp, Susan Lang, Antonette Shibani, Kalliopi Benetos, Chris Anson</i>	713–716
A Framework for Assessing Reflective Writing Produced Within the Context of Computer Science Education <i>Huda Alrashidi, Thomas Daniel Ullmann, Samiah Ghounaim and Mike Joy</i>	717–721
Identifying negative language transfer in writing to increase English as a Second Language learners' metalinguistic awareness <i>Leticia Farias Wanderley, Carrie Demmans Epp</i>	722–725
Towards a Taxonomy of Writing Activities <i>Yoram M Kalman, Laura K Allen</i>	726–727
Writing Analytics to Support Integration of Multiple Texts <i>Jovita M. Vytasek, Alexandra Patzak, Philip H. Winne</i>	728–731
Argument component identification and its application in feedback on Dutch essays <i>Liqin Zhang, Howard Spoelstra, Marco Kalz</i>	732–735
Third Workshop on Social and Emotional Learning (SEL): Integrate SEL and Learning Analytics <i>Elle Yuan Wang, Maria Ofelia San Pedro, Jason Way, John Whitmer and Srecko Joksimovic</i>	736–739
Addressing Drop-Out Rates in Higher Education <i>François Bouchet, Vanda Luengo, Geoffray Bonnin, Anne Boyer, Armelle Brun, Mohamed Amine Chatti, Irene-Angelica Chounta, María Jesús Rodríguez-Triana, Kairit Tammet, Agathe Merceron and Petra Sauer</i>	740–743

Accuracy of a Cross-Program Model for Dropout Prediction in Higher Education <i>Kerstin Wagner, Agathe Merceron, Petra Sauer</i>	744–749
From Data to Intervention: Predicting Students At-Risk in a Higher Education Institution <i>Irene-Angelica Chounta, Kaire Uiboleht, Kersti Roosimäe, Margus Pedaste, Aune Valk</i>	750–755
For and by Student Dashboards Design to Address Dropout <i>Benjamin Gras, Armelle Brun and Anne Boyer</i>	756–761
Curriculum Analytics as a Communication Mediator among Stakeholders to Enable the Discussion and Inform Decision-making <i>Liyanachchi Mahesha Harshani De Silva, María Jesús Rodríguez-Triana, Irene-Angelica Chounta, Kairit Tammets, Shashi Kant Shankar</i>	762–764
Learning Analytics of Student-Centered Advisory Systems in the Introductory Phase of Teacher Education <i>Natalie Kiesler</i>	765–767
A Survey of Learners’ Video Viewing Behavior in Blended Learning <i>Mehrasa Alizadeh, Shizuka Shirai, Noriko Takemura, Shogo Terai, Yuta Nakashima, Hajime Nagahara, Haruo Takemura</i>	768–770

Implementing an attrition model at scale

Amelia Brennan, András Nemes, Jeevan Pokhrel, Cameron Doyle,

Vishesh Jain and David McLay

RMIT University, Melbourne, Australia

david.mclay@rmit.edu.au

ABSTRACT: The development and rollout of a predictive model of student attrition at a large university is described. Student data such as demographic information, enrolment choices and educational outcomes are used to train a machine learning algorithm and subsequently assign each student a score representing their predicted likelihood of dropping out of their program, and also departing the university entirely. These scores, along with considerable contextual information, are provided to program managers, most recently in a pilot project impacting 79 programs with over 17,000 student enrolments. We describe the methods used in developing this model, as well as the experience of communicating the outputs to program managers and the aspects that they have found most useful.

Keywords: Students at risk, predictive modelling, attrition.

1 INTRODUCTION

Identifying students who are likely to drop out of their program before completion is desirable at several levels in a university (see for example Tinto 1987; Manrique et al., 2019; Heisserer and Parette, 2002). Doing so early is key to increasing retention rates and reducing the risk of financial penalty both for students and institutions. While larger tertiary institutions typically do have service capacity to support student outcomes, they often rely on targeting students using single variable descriptors (e.g. a specific demographic variable, past failure in a certain number of courses) or particular business logic developed for a specific cohort only. Using advanced modelling techniques to identify students who are at risk of dropping out of university can enable educators and university administrators to provide targeted interventions and personalised support services for struggling students. Particularly relevant for large cohorts, these techniques can enable time-poor educators to quickly filter and identify at-risk students, and to direct their efforts toward supporting a smaller group. However, while there has been considerable research on attrition modelling, the question of how to provide this information back to instructors or administrators in a way that is timely, useful and actionable is often out of scope or overlooked. This work describes a predictive model of student attrition that has been developed at RMIT University in Australia, as part of a pilot program for 79 higher education programs. In particular, it describes the provision of the predictions – referred to as Early Warning Signs (EWS) scores – alongside contextual student data to the program managers (PMs), who are the educators responsible for managing and administering the program, and who provide a level of academic support for students. EWS outputs are provided at four points during the semester, allowing additional enrolment data to be captured in later outputs, and leading to improved model accuracy. The methods and outcomes described here will inform further rollout of the project to all higher education programs at the institution in 2020.

2 EWS PROCESS

Two predictive model groups, for commencing and returning students, are trained on historical data from 2013 to 2017 (two semesters per year), using data from the Student Information System (SIS). Around 60 variables are included before feature selection: demographic (e.g. age, international status, English-speaking background, employment status), academic (e.g. number of units completed in the program, cumulative GPA, past program enrolment actions) and current term (e.g. academic load, number of campuses, number of courses dropped). The target variable – a binary outcome indicating whether or not a student attrited from a program – is defined as whether the student had any enrolment activity (including taking a leave of absence (LOA)) in any program by the end of the next calendar year. Separate models are trained at several points during the semester: at the ends of weeks 0, 3, 5 and 7, where week 0 is prior to classes commencing and week 5 is the last before census date, after which a student incurs financial liability for their courses. Each release is designed to provide a different focus and benefit to the PMs – the output before the semester start allows PMs to prepare for, follow-up on, or accommodate potentially-at-risk students, whether due to poor prior academic performance, return from LOA or identified-risk enrolment activity. Later outputs that include current-semester engagement metrics (such as online activity and assessment submissions) may alert PMs to students who have a good historical academic profile, but who have begun to struggle in the semester. Over 425k data points (indexed to student/program/semester) are used to train the models, with a random 80/20 training-testing split. Finally, the logistic probability of attrition is produced as an EWS score for all students in a given output.

Individual student EWS scores are emailed to PMs in a spreadsheet containing additional student contextual information along with a summary report. The summary report comprises visualisations of aggregated student demographic data and enrolment information, alongside a summary of the at-risk student predictions. The spreadsheet lists all students in the program with their EWS score, basic demographic information and specific relevant academic variables such as cumulative GPA, number of units completed in the program and academic load. Later outputs include a secondary student risk metric derived from the Learning Management System (LMS) submissions dataset.

All PMs receive outputs at the end of weeks 0 and 5, and are able to opt-in to additional outputs in weeks 3 and 7 – both as a means to gauge interest and to avoid increasing their workloads during these periods. In the week following each output, multiple drop-in sessions are run where PMs are invited to discuss their outputs directly with the developers, and given an opportunity to raise concerns, discuss intervention options, and provide feedback.

3 EWS PILOT

Early iterations of the EWS project, developed by another central analytics team at the university, ran in semester 2, 2018 and semester 1, 2019, for 16 and 36 programs respectively. Most recently, in semester 2, 2019, the EWS pilot was rolled out to 79 programs, with attrition risk scores generated for approximately 17,000 students. The most significant predictive variables were found to be current program status (active or inactive), number of course enrolments, GPA, number of courses dropped, number of deferred programs, gender, socioeconomic status (SES), number of LOAs taken, state of residence, employment status, and parental education level. Of the 79 PMs involved, 25 opted to receive the additional outputs in weeks 3 and 7. The drop-in sessions were reasonably well-

attended, with approximately 20 PMs (25%) attending at least one session over the duration of the pilot and providing valuable feedback.

4 KEY CHALLENGES IN IMPLEMENTATION AND DELIVERY

Technically, the development of the EWS project followed a classical data science project workflow, with the initial challenges of data preparation for modelling that involved hypotheses generation, scoping available datasets, chasing access, understanding definitions, and cleaning and transforming the data. Most of the challenges were typical of most machine learning projects, however the definition of the classifier of attrition deserves note. While attrition has a formal government definition, the classifier for a predictive model is ideally required to be purely a function of the features available in the dataset. Attrition events that are the result of inputs not captured in the dataset can introduce noise and/or bias (e.g. family crisis events, medical events, extra-institutional inputs). Additionally, defining the moment that a student drops out of a program is complicated by the near limitless career sequences allowable, such as: multiple LOAs; switching between programs and between program plans; and exiting a program early with a lesser qualification. This process is naturally further hindered by inaccurate/incomplete capture of program activities in the student enrolment database. Even an unambiguous event such as program discontinuation can be assigned one of 40 reasons. Each circumstance needs to be individually assessed as to whether it is designated a positive label in the attrition classifier. Finally, it is also non-trivial to extract historical training data from the SIS at the equivalent date relative to census in each historical semester – for example, if a student dropped a course *at the census date* in semester 1 of 2013, the training data for the models of weeks 0, 3 and 5 should show the student to still be enrolled in the course.

Another challenging task was addressing the ethical considerations around sensitive and demographic student variables used as features in the predictive model. Sensitive fields such as SES were analysed during the data exploration phase to ensure transparency around model development, understand ethical considerations and ensure no model bias was introduced. Only SES had any marked importance to the classifier and so other sensitive fields were dropped from feature selection. For the final outputs, minimalism, ethics and privacy considerations determined which variables were provided at the level of the individual, and which were included only in the summary report as a program-level aggregate.

However, from an overall project standpoint, the primary challenge was in developing a clear channel of communication with PMs and receiving their support. This required allaying the PMs' concerns early on in discussions and defusing their understandably defensive reaction towards receiving what the executive advertised as an 'at-risk score' of their program's student cohort. The PMs were selected by central administration to participate in the pilot and, based on prior experience, most feared they had been nominated for yet another KPI-driven project accompanied by additional workload and scrutiny. As owners of the project content and delivery, we established consistent messaging early on to the PMs that the outputs were solely to provide them with an additional tool to assist their existing efforts to support students. We ensured this was upheld and demonstrated by listening to their feedback and modifying the outputs correspondingly with each iteration, accompanied by emails with transparent changelogs and invitations for further comments. We ensured the communications contained no suggestions, mandates or calls-to-action based on the data, and instead used drop-in sessions with PMs as an opportunity to discuss and analyse their

data and to obtain a better understanding of their workflow. This provided valuable feedback to the project, which will further increase the benefit to PMs in future iterations. One common criticism from PMs was that the data outputs alone fell far short of providing them with the assistance they need to support their students, and there was a strong suggestion that the pilot expenditure could have been used for providing additional student support resources.

A final difficulty has been in managing the expectations of university executives around a pilot such as this – to avoid a tendency towards overselling, but also to distinguish such a long-term project with the many technicalities that are required to build a robust model from short-term vendor-led projects using LMS clickstream counts that exaggerate their ability to identify at-risk students. We have always tried to be clear about the accuracy and limitations of our model, and pushed back against the idea that such a model should be used as anything other than an additional resource for supporting current students.

5 NOTES FOR PRACTICE

Based on conversations with PMs in our drop-in sessions, the most well-received aspect of the project by the PM stakeholders has been the student-level contextual information provided alongside EWS scores: the demographic breakdown of the student cohort, spreadsheet columns on commencing/returning status, academic load and how many units students have already completed. The predicted EWS score often received surprisingly little attention from the PMs, as for most of them a single source of up-to-date and relevant information regarding their student cohort proved difficult to obtain elsewhere. However, we expect that as the format becomes more familiar, PMs will begin to ask deeper questions of the output including the EWS score. It is therefore most important to provide instructors with understandable data about their students as a path for them to interpret the EWS score in a useful and meaningful way.

When initially delivering a project to educators, we recommend that visible effort is made to ensure that outputs do not add significantly to their workloads. Many staff were initially apprehensive about being nominated as project participants; however most found it both refreshing and relieving to find that the project's aims had no expectations of them built in, but conversely tried to treat them as key customers whose requirements were driving the project.

Finally, it is critical that any provision of at-risk student data to educators be aligned with existing support structures and resources; even the best predictor of attrition is meaningless without the tools to act upon it.

REFERENCES

- Heisserer, D. L., & Parette, P. (2002). Advising at-risk students in college and university settings. *College Student Journal*, 36(1), 69-84.
- Manrique, R. et al. (2019). An Analysis of Student Representation, Representative Features and Classification Algorithms to Predict Degree Dropout. *Proceedings of the 9th International Conference on Learning Analytics and Knowledge*, 401-410, ACM.
- Tinto, V. (1987). *Leaving college: Rethinking the causes and cures of student attrition*. University of Chicago Press, 5801 S. Ellis Avenue, Chicago, IL 60637.

Effects of In-class and Out-of-class Learning Behaviors on Learning Performance and Self-regulated Learning Awareness

Li Chen

Graduate School of Human-Environment Studies, Kyushu University
chenli@mark-lab.net

Yoshiko Goda

Research Center for Instructional Systems, Kumamoto University
ygoda@kumamoto-u.ac.jp

Atsushi Shimada

Faculty of Information Science and Electrical Engineering, Kyushu University
atsushi@limu.ait.kyushu-u.ac.jp

Masanori Yamada

Faculty of Arts and Science, Kyushu University
mark@mark-lab.net

ABSTRACT: This study was designed to investigate effects of different learning behaviors in and out of class on students' learning performance and self-regulated learning (SRL) awareness improvement. The study was conducted at an eight-week information technology course for 70 university students. Results revealed that during in-class activities, using functional tools such as markers and annotations with the support from instructors would benefit learning performance and SRL awareness. Out-of-class activities focusing on specific pages showed positive effects on learning performance and SRL awareness.

Keywords: Learning analytics, Self-regulated learning, In-class behavior, Out-of-class behavior

1 INTRODUCTION

As a critical component of self-directed strategic learning, SRL is closely related to cognitive and metacognitive learning processes and strategies (Pintrich and De Groot, 1990). Several difficulties have been experienced in conducting the two major methods to measure students' SRL processes and activities, which are observation and self-reports. It has been found that in-class observation is labor-consuming and lacks efficiency, while self-report responses rely on students' self-perception, which may cause a lack of accuracy (Tobias and Everson, 2009; Winne and Nesbit, 2009). In this regard, the Learning Analytics (LA) approach shows great promises in data collection and analyses for various situations, especially certain unobservable learning situations. For example, learning logs can automatically collect both in-class and out-of-class learning behaviors that are unable to be observed. These unobservable behaviors include preview and review learning activities without support from instructors. Therefore, this study adopted the LA approach to explore the effect of different learning behaviors in and out of class on learning performance and SRL awareness.

2 METHODOLOGY

2.1 Design Overview

The study was conducted at an information technology course during eight weeks (one 90-minute lecture per week) for 70 freshman students at a university in Japan. Before each lecture, the instructor uploaded learning materials on the BookRoll system (e-learning material reader system) and asked students to preview the learning materials by using some functional tools such as Marker function (highlighting content that they considered important or did not understand), Annotation function (adding annotations to the content), Bookmark function (posting bookmarks), and Search function (searching content by keywords) (Ogata et al. 2017). During the lecture, students were also required to access learning materials on their own devices, add markers they considered important, delete markers or annotations after the reconsidered and further understood the current contents.

2.2 Data Collection and Analysis

First, after each lecture, students were required to spend approximately 10 minutes on a test to assess their comprehension of the current lecture. The final scores (100 full marks) were calculated by the sum scores of eight tests (80 full marks) and assessment for assignments (20 full marks), which represented their learning performance on the course. Second, before and after the course, the pre-post Motivated Strategies and Learning Questionnaire (MSLQ) developed by Pintrich and DeGroot (1990) was implemented to check students' change in SRL awareness. In addition to the five factors proposed by Pintrich and DeGroot—self-efficacy, intrinsic value, cognitive strategies, self-regulation, and test anxiety—an additional factor, help seeking (Wolters et al., 2005), was also added to the pre-post questionnaires in order to assess how students sought help and support during the course. There were 47 items (44 items were about factors from MSLQ, 3 items were about help seeking factor) that were rated on a 7-point Likert scale (1: Strongly disagree-7: Strongly agree). The sum of 47 items was calculated to represent the SRL awareness.

Third, 13 types of learning logs which represented learning behaviors on operating digital learning materials were collected and analyzed. These learning logs were specifically: *Next/Prev*: turning to the next/previous page, *JP*: jumping to a certain page, *Search*: search for keywords, *CL*: click the provided link, *AM/DM*: add/delete marker, *AB/DB*: add/delete bookmark, *JB*: jump to an added bookmark, *AA/DA/CA*: add/delete/change annotation. In this study, we calculated 13 types of learning logs in and out of class respectively, which represented students' learning behaviors during lectures (in-class) and preview and review behaviors (out of class), and analyzed effects of learning behaviors during different learning situations on students' learning performances and SRL awareness. In our study, we tried to build a model for predicting learning performance and improvement of SRL awareness during the different learning situations, thus we conducted stepwise multiple regression analysis to identify significant predictor variables of learning behaviors.

3 RESULTS AND DISCUSSION

3.1 Effects of Learning Behaviors In and Out of Class on Final Scores

First, we used final scores as dependent variables, and 13 types of learning behaviors in and out of class were set as independent variables. As for the results, when the final score was set as the

dependent variable, independent variables such as *Prev_in* ($\beta = .273$, $p = .023$), *JP_in* ($\beta = .306$, $p = .050$), *AA_in* ($\beta = .193$, $p = .094$) of in-class behaviors, and *Next_out* ($\beta = .334$, $p = .063$), *JP_out* ($\beta = -.546$, $p = .007$), *JB_out* ($\beta = .243$, $p = .056$) of out-of-class behaviors accounted for 15.3% variance in final scores. The results were presented in Table 1. From the regression coefficients, it was indicated that during the lecture time, the increased frequency in behaviors of turning to the next page, jumping to a certain page, or adding an annotation, positively affect students' final scores. The same tendency of turning to next page and jumping to a certain bookmark occurred in out-of-class behaviors, were also the predictors for increase in final scores. However, when students jumped to certain pages more frequently, their test scores were more likely to decrease.

Table 1: Results of regression analysis predicting final scores with learning behaviors

Variable	B	SE B	β	t
Prev_in	0.005	0.002	0.273	2.329*
JP_in	0.029	0.014	0.306	1.995 [†]
AA_in	0.007	0.004	0.193	1.700 [†]
Next_out	0.006	0.003	0.334	1.893 [†]
JP_out	-0.241	0.087	-0.546	-2.784**
JB_out	0.262	0.134	0.243	1.950 [†]

$R^2 = .227$, Adjusted $R^2 = .153$, ** $p < .01$, * $p < .05$, [†] $p < .1$

3.2 Effects of Learning behaviors In and Out of Class on SRL Awareness

When difference in pre-post MSLQ was set as the dependent variable, learning behavior variables such as *CL_in* ($\beta = -.274$, $p = .067$), *AM_in* ($\beta = .194$, $p = .085$) during the lecture time, *Next_out* ($\beta = .708$, $p = .001$), *Prev_out* ($\beta = -.616$, $p = .004$), *DA_out* ($\beta = .462$, $p = .002$), *JB_out* ($\beta = .237$, $p = .034$) accounted for 22.6% of variance in the dependent variable of MSLQ change. The results are presented in Table 2. In the lecture, increased frequency of behaviors in adding markers was shown to promote students' SRL awareness. Conversely, when students clicked the provided links more frequently, their SRL awareness was inclined to be negatively affected by this behavior. During students' out of class activities, such as preview or review learning, behaviors including turning to the next page, deleting an annotation, and jumping to certain bookmarks were shown to positively affect students' SRL awareness, while behaviors such as turning to previous pages showed negative effects on SRL awareness.

Table 2: Results of regression analysis predicting SRL awareness changes with learning behaviors

Variable	B	SE B	β	t
CL_in	-10.041	5.390	-0.274	-1.863 [†]
AM_in	0.039	0.022	0.194	1.749 [†]
Next_out	0.074	0.022	0.708	3.401**
Prev_out	-0.204	0.068	-0.616	-3.009**
DA_out	122.077	37.533	0.462	3.253**
JB_out	1.493	0.689	0.237	2.167*

$R^2 = .293$, Adjusted $R^2 = .226$, ** $p < .01$, * $p < .05$, [†] $p < .1$

4 CONCLUSION

This study examined effects of different learning behaviors in and out of class on university students' final scores and SRL awareness improvement. According to Shimada et al. (2015), it is indicated that students' learning performance can benefit from out-of-class learning activities, and results in this study also showed the same effects of out-of-class learning on SRL awareness. Results indicated that students who used certain functional tools such as marker and annotation more frequently during in-class activities tended to perform better in learning performance and improved SRL awareness. Therefore, students are encouraged to use functional tools during the lecture, especially they are provided with the support or instructions about how to understand the contents. However, access to provided reference or links without guidance or instructions was shown to negatively affect the SRL awareness. While during out-of-class activities such as preview work, rather than the utilization of functional tools, focusing on specific pages when reading learning materials showed positive effects on learning performance and SRL awareness, because students were aware of what information they want to know or they are looking for. Thus, students are encouraged to create some bookmarks or save the pages they think important and focus on these pages when conducting individual learning. However, jumping to a certain page without purpose (for example no bookmarks on this pages), negatively affected the learning performance. In addition to learning behaviors reading digital learning materials, future studies must consider other learning behaviors.

ACKNOWLEDGMENT

This research is supported by a JST AIP Grant No. JPMJCR19U1, Japan.

REFERENCES

- H Ogata, Y Taniguchi, D Suehiro, A Shimada, M Oi, F Okubo, M Yamada, and K Kojima. (2017). M2B System: A Digital Learning Platform for Traditional Classrooms in University. Proceedings of LAK2017, 155–162.
- PR Pintrich and EV De Groot, (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology*, 82(1), 33.
- A Shimada, F Okubo, C Yin, K Kojima, M Yamada, H Ogata. (2015). Informal learning behavior analysis using action logs and slide features in e-textbooks, Proceedings of IEEE ICALT 2015, 116–117.
- S Tobias and HT Everson. (2009). The importance of knowing what you know: A knowledge monitoring framework for studying metacognition in education, *Handbook of Metacognition in Education*. Routledge, NY, NY, USA.
- PH Winne and JC Nesbit. (2009). Supporting Self-Regulated Learning with Cognitive Tools. *Handbook of Metacognition in Education*, Routledge, New York, NY, USA.
- CA Wolters, PR Pintrich, and SA Karabenick. (2005). Assessing academic self-regulated learning. In *What do children need to flourish?* Springer, Boston, MA, USA.

Learning Analytics Applied in Commercial Aviation: Our Use Cases, Obstacles, and Help Needed From Academia

Liz Gehr, Ph.D., Laurie Dunagan

The Boeing Company

Liz.Gehr@boeing.com; Laurie.Dunagan@boeing.com

ABSTRACT: The growing demand for future aviation workers, including pilots, maintainers, and cabin crew, along with the need to maintain the highest quality and safety standards, results in a need to rethink our traditional approach to aviation training. Aviation training has typically followed the traditional model of classroom training, computer based training, simulator based training (for pilots) and hands on training (for mechanics). At the same time, the regulatory authorities are moving to a competency-based training method for measuring and certifying pilot readiness, rather than the traditional time-based criteria. This practitioner presentation will describe obstacles we face in our attempt to ameliorate some of these issues presented by these two realities, through adoption of a training analytics program in aviation training. It will outline the strides we have made, the challenges we face, and how we believe the academic learning analytics community can work with us.

Keywords: Aviation; Competency Based Training

1 BACKGROUND – WHY LEARNING ANALYTICS IN AVIATION

The current forecast for new civil aviation pilots, worldwide, projects that to meet future demand an additional 804,000 pilots, 769,000 maintenance technicians, and 914,000 cabin crew will be needed in the next 20 years (Boeing, 2019). This will require a radical shift for industry to recruit and train personnel to meet this demand. In fact, we want to do this while sustaining a long-term trend of continuous improvement in safety and quality across the industry. Therefore, we are analyzing ways to achieve these two objectives: increase trainee output while improving the educational experience for a well-trained aviation professional. Some new methods being looked at include using Augmented Reality (AR), Virtual Reality (VR), and Virtual Instructors (VI) to better prepare students before they train in a Full Flight Simulator (FFS). FFS time is a finite and expensive resource, and can be used more effectively and efficiently. Separately, some airline regulators are requiring a transition to competency-based methods and metrics to train and assess pilots from the historic method of time-based training (EASA 2015).

One way the industry is considering meeting the requirements of increasing pilot demand and competency-based learning is through the application of a robust learning analytics program. Such a program would allow us to better assess what knowledge students have mastered and when, so we can tailor training to the student, providing the right course, at the right time, to the right student, in the right fashion. In effect, we would be personalizing learning for each student, focusing on and addressing areas where each individual trainee has knowledge gaps rather than subjecting all students to the same training, as is often the case now.

Learning analytics would also advantageously support competency-based training. Since competency-based learning or training focuses on the degree of mastery and not just providing training, it ensures that the learner attains the needed level of learning or competency before moving on to the next level of learning. This personalizes learning; in some cases training times can be shorter when the learner achieves mastery of the competencies quickly, or longer, when a learner needs more time to master a skill.

In addition, we believe learning analytics could help us identify when a student has achieved competency, as well as measure and locate learning gaps, much earlier than we are able to now. For example, learning analytics may be able to identify a potential deficiency that may not otherwise be observed and alert a training manager for further evaluation or additional study.

The goal of a competency-based training approach coupled with learning analytics is to match the training to the learner and assess learning outcomes. The learner builds on prior knowledge and experience to progress, and competency-based learning lessens or even eliminates skill gaps that may occur in traditional learning methods. Learning analytics provides evidence for competency assessments, by looking at learning activities and obtaining relevant metrics. Measurable behavioral objectives are specified within the learning activities, and data collected to conduct learning analytics. These data elements may come from lectures, reading, scoring rubrics, quizzes, learner engagement, performance, etc. One need we have in order for competency-based training to be more effective, is better assessment tools and methods for evaluation of the competencies, especially when looking at soft competencies.

The need to measure and assess these soft competencies is a need in the aviation training industry. As opposed to many skills that the learning analytics community has examined such as math, physics, or foreign language mastery, many skills that pilots must master are more difficult to measure – for example, communication, leadership, and situational awareness. These, too, are essential skills that are currently trained and evaluated, but moving these to a competency-based training program based on learning analytics presents unique challenges.

The remainder of this paper will outline the unique obstacles the aviation industry faces in attempting to implement a robust learning analytics program, and put out a request for help from the learning analytics industry.

2 OUR PREVIOUS EXPERIENCE WITH LEARNING ANALYTICS

As a company, we have developed an analytics framework based on an existing military training database. Although the database was designed to track training, not to provide analytics, we were able to look for trends in the data and identify opportunities for improvement in the training provided. We then used this experience to design a framework for large scale learning analytics in military or commercial aviation. In conjunction with learning analytics, we are also designing competency-based training as a way to enhance our analytics framework. However, during implementation, we are running into many obstacles unique to our environment when trying to use this framework in either the military or commercial domain.

3 OUR OBSTACLES

Although many in the aviation industry have been following the advances made in learning analytics research and development, attempts to fully implement a learning analytics program in commercial aviation face many unique challenges. One obstacle not unique to aviation is privacy. There is a moral, as well as legal, need to protect the data of all trainees in an aviation learning analytics program. The data must be protected from being accessed by inappropriate or unauthorized individuals, either by accidentally allowing, for example, a co-worker to see a fellow student's performance, or from hacking from an actor with malicious intent. In addition, many countries follow the European Union's (EU) General Data Protection Regulation (GDPR) which gives people many rights, and places additional responsibilities onto data collectors. These include the right to access one's own data, the right to have their data erased, and the right to restrict automated decision making based on their data.

All learning analytics implementations face these basic privacy restrictions, but the aviation industry has additional, unique challenges. Many aviation professionals are in a union, which provide additional rights to their members. Also, due to the global nature of the airline industry, and the mobility that aviation professionals have due to their unique skill sets, we can't look to the laws of one country, or one region, to guide our behavior when it comes to collecting, storing, analyzing, and using data. This adds additional complexity and responsibilities for the data collector. It becomes obvious quickly that to implement a learning analytics program, data privacy and personal privacy rights must be addressed at the outset.

Another area where the aviation industry has unique struggles is accessing all the data that would be needed for a well-implemented learning analytics program. Students come to us with very diverse backgrounds and a wide breadth of aviation knowledge. Some may be just starting their career in commercial aviation, having earned their hours in private aviation, flying small planes, and need to be trained in the basics of handling a large jet. Other pilots come to the commercial aviation world after retiring from the military where they flew thousands of hours in a cargo, fighter, or bomber aircraft. Currently, students can fill out a demographic questionnaire, but this doesn't capture the true richness and diversity of their past experiences. In one military database we examined, where the demographic form was not required, only 20% of students provided any of this type of data. This is not enough information to conduct a thorough analysis. In addition, past training records would be spread across multiple systems with varying degrees of completeness and accessibility. To properly perform learning analytics, we would want to know the history of each of these students.

While some data collection is relatively easy - gathering data from computer based training, or instructor ratings of students, for example - obtaining other useful training data is much more difficult. Data from full flight simulators is difficult to use as there are more than 30 manufacturers of these simulators, and there are no standards for data output. For example, knowing the average altitude for a portion of a flight would not be in the same data format in two different simulators. This is in contrast to military simulators which have communication standards making data extraction, although still not simple, at least easier.

In addition, the gold standard of training analytics data is an analysis of student job performance. The same model of plane can be flown by over a hundred airlines, and in the case of the 737, a

variant is also flown by the military as the P8. Ideally, we would be able to combine data – both training and operational – from pilots from different airlines to form a larger database and draw more robust conclusions. However, while data is recorded for every commercial flight, obtaining, analyzing, and most importantly, tying the data to an individual is filled with challenges. Additionally, obtaining information pertaining to pilot performance from airlines or the military after their initial training is completed is very difficult, as many do not want to share this type of information. Customer agreements vary by each airline, and legal issues may prevent us from combining data to form a rich database of information.

4 OUR REQUEST FOR THE LEARNING ANALYTICS COMMUNITY

As noted, we have many obstacles to overcome before we can implement a full learning analytics solution. Even if the aviation industry overcame these challenges, it does not have the resources, or ability, to implement a “gold standard” effectiveness study, with some students participating in a learning analytics based training program, while others go through a standard program, with a goal of demonstrating that the learning analytics students perform at least as well as the traditional students. Instead, we are in a position of having to prove that a new method of training is effective, before we can even implement it. For obvious reasons, we are a risk averse and highly regulated industry and need to know that what we implement will improve our outcomes before we make a change.

Therefore, we wish to propose a collaboration with the broader learning analytics community, to determine how we can work together to both enrich the base of knowledge about learning analytics, as well as gain insights to help us prove to the regulators that implementing a learning analytics program in commercial aviation will lead to improved training. We believe all parties would benefit from such a relationship. The broader academic learning analytics community would have access to a real world use case that may validate their theories, and the aviation industry would benefit by having a solid theoretical foundation for their learning analytics implementation.

Ideally, the outcomes of this collaboration would include evidence of previous successful implementations of large scale learning analytics programs; this evidence could be shared with customers to help them build trust in implementing learning analytics. Successful examples of competency-based learning, and evidence that personalized adaptive learning works, are necessary before it can be implemented. By collaborating with the academic community, we believe we can work with the regulators, our customers, and the larger aviation training community to implement a successful learning analytics program that maintains, or even increases, the competence level of aviation personnel.

REFERENCES

- The Boeing Company, 2019. Retrieved September 9, 2019, from <http://www.boeing.com/commercial/market/pilot-technician-outlook>
- EASA, 2015. ToR RMT.0696 Implementation of Evidence-Based Training within the European regulatory framework.

For Evidence-Based Class Design with Learning Analytics: A Proposal of Preliminary Practice Flow Model in High School

Satomi Hamada¹, Yufan Xu¹, Xuewang Geng¹, Li Chen¹,
Hiroaki Ogata², Atsushi Shimada³, Masanori Yamada^{1,4}

Graduate School of Human-Environment Studies, Kyushu University¹

Academic Center for Computing and Media Studies, Kyoto University²

Faculty of Information Science and Electrical Engineering, Kyushu University³

Faculty of Arts and Science, Kyushu University⁴

satomi.hamada@mark-lab.net, xuyufan@mark-lab.net, geng@mark-lab.net, chenli@mark-lab.net, hiroaki.ogata@gmail.com, atsushi@ait.kyushu-u.ac.jp, mark@mark-lab.net

ABSTRACT: This paper introduces a practice to incorporate a learning analytics dashboard that analyzes and visualizes learning logs using digital textbooks for high school students. Based on the knowledge gained through the practice over the past six months, an important model for incorporating learning analytics in high schools is proposed. In the future, we plan to examine the model, including how learning logs can be used for each teacher's class.

Keywords: Learning Analytics, Study Logs, Digital Textbooks, Proposal Models, High School

1 INTRODUCTION

As a promising field, learning analytics (LA) is widely applied to understand and improve the learning environment, which involves educational data collection and analysis. With the support of technology, it enables the collection of educational data from various resources, for example, learners' profiles (such as their education level and motivation), curriculum profiles (such as information of curriculum and learning materials), and learning logs collected from online systems or learning material reader system that identifies online learning behaviors, such as time students spent on each activity or the operations on reading learning materials (Ifenthaler and Widanapathirana 2014; Yamada, 2017). To utilize LA approach to improve the learning environment, Dunbar et al. (2014) indicated that curriculum designers play an important role in bridging data collection and analysis with instructional design considering students' learning conditions. The primary/secondary education involves various stakeholders in addition to students and teachers, and their decision-making needs to be included for effective and efficient LA-based practice. This paper proposes a preliminary work flow model of the practical development of LA for decision-making.

2 CASE STUDIES

This practical research project is a collaborative project that involves three university laboratories (one lab: educational technology, other labs: information science), a public high school, and local government. This practice was conducted on eighty 10th grade students. Each student is provided with a tablet device, and mathematics (one class/day) and English (two classes/week) classes are conducted. In mathematics, students are divided into three classes by proficiency level, and in

English classes, students are taught together regardless of proficiency.

Table 1: The characteristics of the four teachers

	Teacher A	Teacher B	Teacher C	Teacher D
Subject	Mathematics	Mathematics	Mathematics	English
Teaching Experience	20 years	30 years	8 years	4 years
Teaching style	Mainly using textbooks, Solving questions during preparation and explaining them in class, There are few questions from students, The system is not used much	Mainly using textbooks, An environment where students can easily ask questions, Teacher frequently comments on system usage evaluation Mainly	Using my own prints, An environment where it is difficult for students to ask questions, The system is not used much	Mainly using textbooks, A small test of English words, Student learn in pairs, Speaking in class is not permitted, Regular note check.
System usage	Submit assignments using Moodle, Review based on student markers	Use BookRoll in preparation, Submit assignments using Moodle, Use BookRoll for assignments, Review based on student markers	Submit assignments using Moodle, Review based on student markers, Distribute question paper with iPad's airdrop	Use BookRoll in preparation, Submit assignments using Moodle, Use BookRoll for assignments, Review based on student markers, Conduct a small test with Moodle

2.1 The system used in this practice

There are three kinds of systems used. First, a learning management system, “Moodle”, is used. In this practice, in addition to its role as a portal for the learning system, it is also used as a test plug-in and an assignment submission plug-in. The second is the digital teaching material viewer “BookRoll” (Ogata et al., 2017). “BookRoll” has the ability to mark the displayed digital teaching materials, make notes (typing and handwriting formats), as well as a bookmark function, a search function, and a recommendation information presentation function. The third is the “Dashboard system” (Flanagan and Ogata, 2018). The dashboard displays a graph of the learning logs accumulated by textbooks and uploaded to the “BookRoll”. Specific learning logs include the actions taken by the students on each page (marker count, marker point, number of notes, and note content) and the time spent by them when they were looking at a page. These systems were introduced at the start of this practice.

2.2 The teachers who cooperated in practice

The characteristics of the four teachers who cooperated in the practice in this study are given in Table 1. All students read textbooks using “BookRoll” to write assignments outside the class and clarify their understandings. All teachers instructed them to submit assignments on Moodle. Before a class starts, each teacher views the dashboard, decides the class plan, and then conducts the class.

2.3 The educational design assistant

In this practice, educational design assistants are provided by the university to introduce three systems and other assistance in the high school every weekday. There are five roles of the educational design assistant: 1. To provide information about the three systems to teachers, 2. To create system manuals, 3. To answer any question about system usage for teachers and students, 4. To handle problems during classes when the systems are used; if any problem occurs, then they need to take it back to the university and get in touch with the system developer, and 5. To provide a suggestion regarding instructional design based on the dashboard and learning analytics data, such as the data shown in Table 2. Referring to Table 1 and Table 2, it is suggested that the usage

Table 2: Sample learning logs for educational design

Week 1		FLIP NEXT	FLIP PREV	ADD MK	ADD MEMO	ADD BM	DELETE MK	DELETE MEMO	DELETE BM
In class	M Class A	546	185	468	21	12	353	8	10
	M Class B	975	354	244	19	9	123	2	8
	M Class C	172	15	20	45	0	1	2	0
	E Class 1	300	102	283	11	2	288	11	2
	E Class 2	94	35	275	8	0	213	0	0
Out-of-class	M Class A	575	96	480	8	1	137	2	1
	M Class B	387	80	128	17	2	39	0	2
	M Class C	749	143	276	0	1	93	0	1
	E Class 1	652	189	644	18	3	371	12	3
	E Class 2	397	80	509	5	2	300	0	1

Note: M Class: Mathematics Class, E Class: English Class, MK: MARKER, BM: Bookmark

frequency and usage method of the system varies greatly depending on the teacher, and the view of education also differs. After all classes are complete, educational assistants discuss the problems of class with teachers and sometimes the head of the school, in order to improve class design and intervention during class. In the future, educational assistants will mainly implement interventions to develop teachers' ability based on analysis tools, such as "Dashboard", rather than providing support during class.

2.4 The workshop

In this practice, learning analytics workshops were held. Teachers, educational design assistants, other stakeholders, such as the head of the school, and the municipal board of education attended these workshops. In these workshops, issues and problems arising from the practice will be considered and discussed together with the head of the high school and the municipal board of education to design an effective curriculum and methods for instruction and how this can be used in other schools. For example, participants, based on the data as displayed on the dashboard, spoke about the educational practices that uses learning analytics and discussed effective seamless learning design that bridges in- and out-of-class practices using the three systems. To make the practice with learning analytics more effective, teachers suggested the educational technology and information technology researchers to add new functions to these three systems and the municipal board of education to improve the viewer and dashboard. The head of the school and municipal board members also asked other stakeholders to proceed with the evidence-based instruction and analyze the effects of these instruction using learning analytics. These workshops were conducted once a month to implement an effective organizational learning analytics. This workshop seems to contribute to enhancement of teachers' motivation for the practice, and new instructional design with LA platform were proposed by sharing the concerns of each school. This Workshop activities beyond their region are encouraged.

3 PRELIMINARY FLOW MODEL FOR EVIDENCE-BASED CLASS DESIGN

Based on the practice so far, it will introduce an ideal model for implementing class designs using learning analytics by introducing the system in a school (Figure 1). There are two important points in this model. First, activities are performed at different granularities, and the research team including educational design assistants manages activities at each granularity. Second, there is interaction among teachers who are using the system. The practice is promoted by sharing ways to incorporate

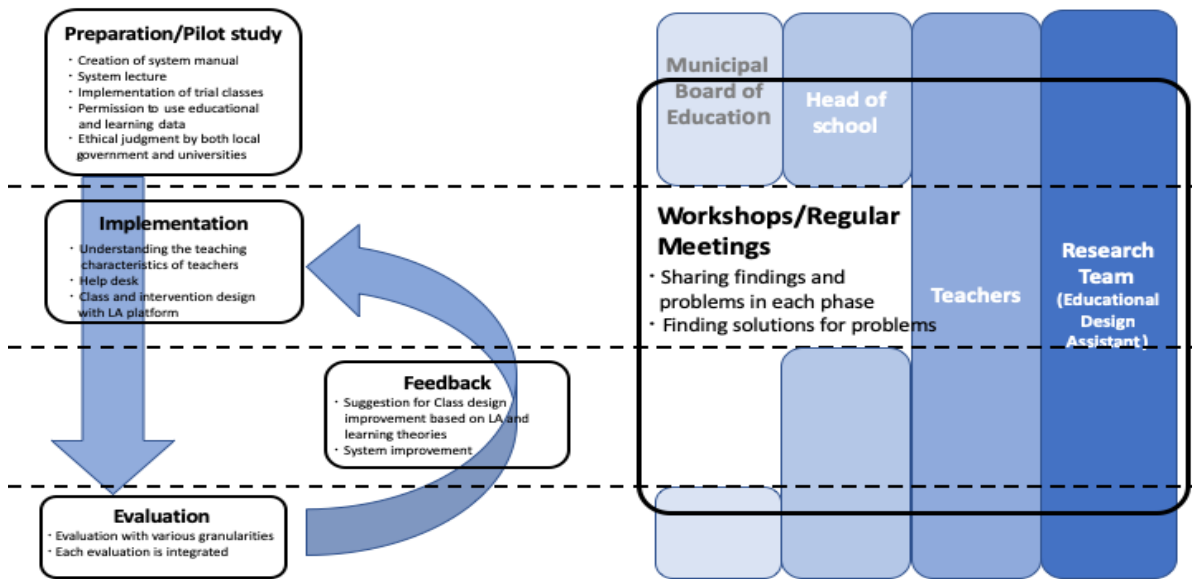


Figure 1: Preliminary implementation flow model for evidence-based class design in high school

the system into lessons and raising concerns regarding the system. This practice provides the opportunity to share information with practice schools in other prefectures. Using these efforts, it could be possible that each organization can cooperate and practice smoothly.

4 CONCLUSION AND FUTURE WORK

In this paper, a flow model for evidence-based class design in high schools was considered based on actual practical experience. As future works, the project members are required to discuss the important elements, data and events through workshop, discussion, observation, intervention, and reflection with data through formative evaluation, in order to modify the proposed preliminary flow model. The negative and positive impact and effects of the feedback on stakeholders' awareness in terms of education, project team, and policy should be investigated.

ACKNOWLEDGMENT

We would like to thank everyone at F high school for their cooperation in this research. This research is supported by JST AIP Grant No. JPMJCR19U1, Japan.

REFERENCES

- Flanagan, B. & Ogata, H. (2018). Learning analytics platform in higher education in Japan. *Knowledge Management & ELearning*, 10(4), 469-484
- Ifenthaler, D., & Widanapathirana, C. (2014). Development and validation of a Learning Analytics framework: Two case studies using support vector machines. *Technology, Knowledge and Learning*, 19(1-2), 221-240.
- Ogata, H., Taniguchi, Y., Suehiro, D., Shimada, A., Oi, M., Okubo, F., Yamada, M., & Kojima, K. (2017). M2B System: A Digital Learning Platform for Traditional Classrooms in University, Proceedings of LAK2017. 155-162
- Yamada, M., Shimada, A., Okubo, F., Oi, M., Kojima, K. & Ogata, H. (2017). Learning analytics of the relationships among self-regulated learning, learning behaviors, and learning performance. *Research and Practice in Technology Enhanced Learning*, 12, 13.

Building a Unified Data Platform to Support Metacognition and Self-Regulated Learning

John Johnston

The University of Michigan
johnpj@umich.edu

Etienne Pelaprat

Unizin
etienne.pelaprat@unizin.org

ABSTRACT: To capitalize on the promise of personalized learning analytics, our applications must rely on a comprehensive view of student learning in data. This becomes especially true as student-facing analytics evolve from descriptive analytics into experiences that scaffold student reflection on their own learning. Ironically, as the number of tools used by students expands such that data becomes abundant, so the data ecosystem become fragmented, challenging our ability to build valuable analytics. In short, to build personalized analytics you must first solve a key problem: from many tools, build a comprehensive, unified portrait of the learner in their learning environment. With such a foundation in place, it becomes possible, we believe, to build effective, analytics-driven student experiences. In this project, we report on an effort to build a unified data ecosystem and a personalized learning application in parallel.

Keywords: learning analytics, personalized learning, metacognition, data ecosystem

1 INTRODUCTION

Over the last decade, the academy and its technology suppliers have made great strides in developing learner analytics that enable student experiences to foster reflection, metacognition, and self-regulated learning (Durall & Gros, 2014). Most successful examples of such analytics, as in tutoring systems and adaptive learning platforms, rely on the fact that they deliver and observe as much of the student's learning experience as possible. Behind these tools' success is an important principle: the more comprehensive your view of the learner and their behavior in the context of the course design, the richer the analytics-driven experience can be. Indeed, we know that comprehensive representations in data of learners, learning processes, and learning environments are required to capitalize on the promise of personalized education at scale through research-based learning analytics (Bingham, Pane, Steiner, & Hamilton, 2018).

Beyond comprehensive individual learning tools, however, the teaching and learning data landscape is highly fragmented and heterogeneous (Bodily & Verbert, 2017). Most course experiences still rely on a handful of tools to design and deliver student experiences. Their individual tool data is not easily made available to other tools, if at all. Consequently, while dozens of learning analytics features and tools for students exist, all are limited to using the data from the tools in which they are embedded.

Our project – to build a unified learning data platform as the foundation for student-facing analytics – realizes a solution to the two key problems hindering advancement in personalized, data-informed learning experiences. First, it creates a more comprehensive and consistent representation of learners, learning and environments in data. Second, it demonstrates the generation and use of learning analytics based on multiple data sources, integrated into the student's learning experience.

2 UNIFIED DATA PLATFORM AND ANALYTICS TOOLS

Our unified learning data platform leverages cloud-native infrastructure to ingest and commonly model teaching and learning data from across the tool ecosystem into a common set of data services. The platform integrates two forms of data generated or emitted by teaching and learning tools: 1) behavioral data in the form of real-time event streams; and 2) contextual data generated out of an operational data store. By transforming all behavioral, performance, and demographic data from each tool into a common data model, the platform creates a standard foundation for data services that offer a 360-degree view of learners, learning processes, and learning environments. Today, that platform is being piloted by 9 large US research institutions, captures millions of records of data, and offers a handful of common data services. This enables those institutions to develop common services on common data infrastructure to build solutions to shared problems.

Riding atop this platform is a student-facing dashboard which serves to support metacognition and self-regulated learning through a suite of thoughtfully tailored visualizations. This dashboard ingests learning event streams (created from multiple tool sources) and joins them with performance data to provide the students with visualizations that reflect their participation and performance in the course and in relation to their peers. For instance, students are provided a ranked view of the most popular course content and can then filter that content by performance level (e.g. students scoring 90-100%, 80-89%, etc.). Any content that I as the student have not viewed, will be highlighted to draw my attention [figure 1]. Content I have viewed will display details such as the last time I opened it and how many times I've viewed it. There is a direct link to the files from the visualization, so I can quickly review popular content I may have missed. In this way, our design focuses on mastery of content and learning goals and not just grade performance. Other capabilities include a learning activity timeline, assignment planning tool [figure. 2], and grade distribution visualization.

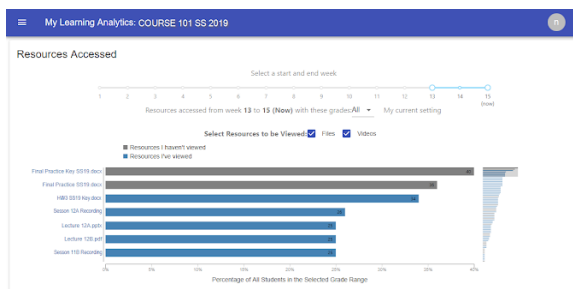


Figure 1: Resources Accessed

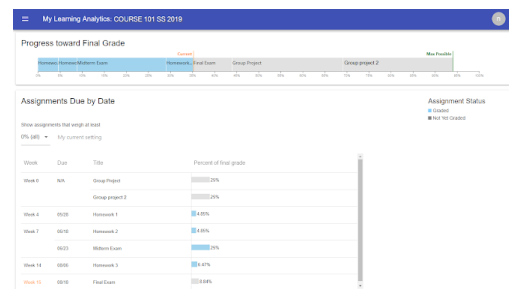


Figure 2: Assignment Planning

The design of the visualizations was informed by the emerging body of research into the efficacy of student-facing data dashboards, self-regulated learning (SRL) theory, and the observed shortcomings with past learning analytics products (Tsai & Gasevic, 2015). A key design principle of this dashboard is that visualizations are presented inline with the students' normal learning workflows, where they are most impactful. In many ways, this approach aims to remedy to what research (Wise, Vytasek, Hausknecht, & Zhao, 2016) has suggested is the flawed approach of overwhelming students with data dashboards that are difficult to interpret and don't provide a clear path to improving learning behaviors.

3 IMPLEMENTATION PROCESS

We conducted our research study on the use of this dashboard to determine if students perceive that these visualizations have impacted their motivation, self-regulated learning, or overall course performance. A set of three visualizations were made available in the learning management system (LMS). Students could access the main dashboard through a link in the course site navigation or use

one of the context-specific links. The context-specific links provided direct access to the visualization in the LMS tools where they were most relevant: 1) The Files Accessed visualization in the Files tool; 2) the Assignment Planning visualization in the Assignment tool; 3) and the Grade Distribution visualization in the Gradebook tool. In this way, the insights were as close as possible to the student's normal learning workflow. The students were provided a 15-minute orientation to the tool that included a demonstration of the basic functionality.

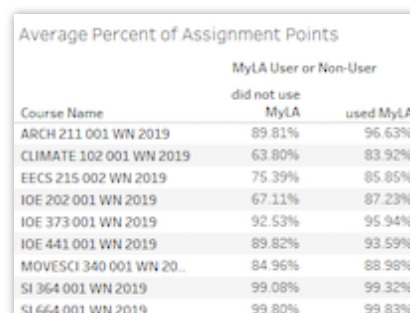
We have piloted the system in two academic terms. The first term ran from September 4th to December 31, 2018 and included 3 pilot courses with 336 students participating. The second term ran from January 9th to May 2, 2019 and included 10 courses with 449 students. Each pilot course participated in a pre and post survey. The log data was also analyzed to surface viewing behaviors.

4 RESULTS AND FINDINGS

In our initial analysis of term 1 student engagement (n=178 responses) with these visualizations, 88% of students agreed it changed the way they plan their course activity; 87% of students reported that use of the visualizations increased their sense of control over their course performance; and 84% agreed it improved their performance in the course. By “sense of control over their course performance” we sought to measure whether the tool increased their sense of agency over their grade outcome. As one student responded, “Periodically, I used it to check up on where I was in the class, understand things I had to get done, and gauge my overall understanding in order to effectively roadmap longer term goals.”

The log data showed that 82% of all students viewed the dashboard at least once, with an average viewing time of almost 2 minutes per page. The Files Accessed visualization had the most views (760), followed by the Assignment Planning view (475), and the Grade distribution being the least viewed (201). It is important to note that one of our three courses in term 1 did not enable the grade distribution view.

In our second term of the pilot students reported that dashboard use changed their sense of control over their course performance (76%) and changed the way they planned their course activity (68%). The Grade Distribution visualization was the most widely used with at least 2324 views, followed by the Files Accessed view (1578), and the Assignment Planning view (145). In all the courses that participated in the pilot, there was a positive correlation between students that viewed MyLA at least once and that student's performance on graded assignments [Figure 3]. We do not yet have sufficient evidence to suggest that there is a causal relationship, but this does point to an intriguing focus for future research.



Course Name	MyLA User or Non-User	
	did not use MyLA	used MyLA
ARCH 211 001 WN 2019	89.81%	96.63%
CLIMATE 102 001 WN 2019	63.80%	83.92%
EECS 215 002 WN 2019	75.39%	85.85%
IOE 202 001 WN 2019	67.11%	87.23%
IOE 373 001 WN 2019	92.53%	95.94%
IOE 441 001 WN 2019	89.82%	93.59%
MOVESCI 340 001 WN 2019	84.96%	88.98%
SI 364 001 WN 2019	99.08%	99.32%
SI 664 001 WN 2019	99.80%	99.83%

Figure 3: MyLA Use and Assignment Performance

One of the early challenges we encountered was in accounting for the diverse methods of student assessment being employed across the pilot courses. In cases where faculty opted to use less

traditional methods of assessment – such as gamification, mastery-based learning, or forced ranking – the grade distribution visualization may not sufficiently represent the performance of the student. In those cases, faculty may disable that view. We were also very sensitive to inadvertently surfacing information in the grade distribution view that could be used to extrapolate or hint at a given student's performance. We mitigated this risk initially by only including courses with greater than 30 students. We later developed logic to obscure scores the lowest end of the grading scale by binning outliers.

We believe that the student responses regarding the influence of the tool on their metacognitive behavior are strong indicators that we're meeting our intended goals of supporting student agency and in providing learning insights where they are most relevant. As one student explained, "It helped me organize and get a good basis for what I needed to do." To further evaluate the impact of the system, efforts are underway to pilot the student dashboard at an additional six universities in the US and one Canadian university. We have also expanded the types of learning activities that are being represented in the visualizations to include lecture recording and media file viewing with plans to visualize eText views and interactions with student response systems. Future inquiry will be directed towards investigating the apparent correlation between usage and performance.

REFERENCES

- Bingham, A. J., Pane, J. F., Steiner, E. D., & Hamilton, L. S. (2018). Ahead of the Curve: Implementation Challenges in Personalized Learning School Models. *Educational Policy*, 32(3), 454–489. <https://doi.org/10.1177/0895904816637688>
- Bodily, R., Verbert, K. (2017). Trends and issues in student-facing learning analytics reporting systems research. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 309–318. ACM. <https://doi.org/10.1145/3027385.3027403>
- Durall, E. and Gros, B. (2014). Learning Analytics as a Metacognitive Tool. In *Proceedings of the 6th International Conference on Computer Supported Education - Volume 2: CSEDU*, ISBN 978-989-758-020-8, pages 380-384. <https://doi.org/10.5220/0004933203800384>
- Tsai, Y., and Gasevic, D. (2017). Learning analytics in higher education --- challenges and policies: a review of eight learning analytics policies. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 233-242. ACM. <https://doi.org/10.1145/3027385.3027400>
- Wise, A., Vytasek, J., Hausknecht, S., & Zhao, Y. (2016). Developing Learning Analytics Design Knowledge in the "Middle Space": The Student Tuning Model and Align Design Framework for Learning Analytics Use. *Online Learning*, 20(2). <http://dx.doi.org/10.24059/olj.v20i2.783>

Using the Behaviour Change Wheel for Learning Analytics adoption

Hazel Jones

University of Southern Queensland
w005456@usq.edu.au

ABSTRACT: This paper describes the development and piloting of a novel approach to supporting individual staff to adopt Learning Analytics (LA) to inform and enhance their teaching practice and learning design. The Behaviour Change Wheel (BCW), which provides a pragmatic guide for developing and implementing interventions provides the theoretical framework for this study and is explained in the context of this study. Possibilities for more widespread adoption are discussed. Initial feedback from a small (n=6) pilot study conducted over 20 weeks, suggests that the approach has merit with participants noting increased awareness and use, of LA.

Keywords: Behaviour Change Wheel, change management, faculty engagement, learning analytics.

1 INTRODUCTION

Academic staff are a key group of stakeholders in the use and application of Learning Analytics (LA), who require knowledge, skills, motivations, and time to engage deeply; circumstances exacerbated by lack of access to data in an easily useable format (Bichsel, 2012; Jones, 2019; Klein, Lester, Rangwala & Johrey, 2019). Ongoing support and training need to be provided to encourage increased awareness and uptake, and therefore change in attitude towards the benefits of using LA to inform and enhance teaching practice and learning design. Opportunities to connect with like-minded colleagues are also an important aspect of support (Gunn et al., 2017, Rehrey et al. 2018). This paper outlines two phases of an on-going study developing and trialing an implementation plan to support academic staff to engage with LA, describes the design and testing of a 20-week implementation plan. The plan provided support and advice to individual academics from a Learning Design perspective. The study occurred at a regional Australian university and adopted a change management approach using the Behaviour Change Wheel (BCW) (Michie, Atkins & West, 2014).

2 BEHAVIOUR CHANGE WHEEL

The BCW, as outlined in Figure 1 (Michie et al., 2014) provides a practical guide for developing and implementing interventions and policies to effect behaviour change. The approach originated in health and medical fields, built from a synthesis of 19 behaviour change frameworks. Whilst the guide is widely cited and used in many contexts, there is limited evidence of its use in higher education settings, especially for the adoption of LA. Using the BCW as the theoretical framework for this study is distinctive and an example of how this can be adopted in a variety of contexts. The COM-B Model is the hub of the BCW and details three interlinked components of Capability, Opportunity and Motivation. One or more of these aspects need to change to achieve the target behavior and for this study the behavior investigated is engagement with LA by individual academics. Stage 2 of the BCW involves choosing Intervention functions, which are “broad categories of means by which an intervention can change behaviour” (Michie et al., 2014, p.109). In

Stage 3 of the BCW, two or three Behaviour Change Techniques (BCTs) are chosen for each intervention and the mode of delivery for each of these determined.



Figure 1: The Behaviour Change Wheel (from Michie et al, 2014, p. 18)

3 APPLYING THE BEHAVIOUR CHANGE WHEEL TO THIS STUDY

Findings from the initial data-gathering phase of this study (Jones, 2019) identified the main challenges to staff adopting LA were time constraints, and lack of training and knowledge, indicating these are key areas of opportunities to be included in an implementation strategy. The main areas of support staff identified as needing were accessibility and interpretation of data; training and support in use of LA and to a lesser degree, institutional guidelines on how to use. All aspects of the COM-B model, except physical capabilities, are contributing factors in this institution, with physical opportunity and psychological capability identified as the main components. Some consideration of social opportunity and automatic motivation are also needed, as outlined in Table 1.

Table 1: Relevant COM-B Model components for this study (adapted from Michie et al, 2014, p.63)

COM-B components	What needs to happen for the target behaviour to occur?	Is there a need for change?
Psychological capability	Know what tools are available in the LMS	Deepen knowledge, from simple awareness of tools, to good practice use
Physical opportunity (eg time & resources)	More time, training and support provided	This was identified as the main area of need
Social opportunity (interpersonal influences and cultural norms)	Opportunities for discussing and working with colleagues	Yes, approval of colleagues for teaching practice is important so opportunities for sharing would be beneficial
Automatic motivation (emotional reactions, wants and needs)	Gain satisfaction and positive outcomes from adopting LA	Some changes needed to embed LA into the teaching and learning culture and practice

This study chose four intervention functions (enablement, environmental restructuring, modelling and training) due to their links with physical opportunity and psychological capability, and two-three BCTs for each of these were incorporated into the intervention plan. As an example of this process, Enablement (reducing barriers to increase capability or opportunity) involves action planning and

practical support. In this context, action planning included supporting each participant to determine a specific question to investigate, plan how to use LA to inform this investigation, design and implement an action and evaluate its effectiveness. Practical support was provided for each step through individual and group meetings, and a resource site in the LMS (Moodle).

4 IMPLEMENTING THE PLAN

The implementation plan involved support provided over a 20 week period, covering one semester, including three individual sessions with each participant and three group meetings; learning design support; and a support site in the LMS providing resources and discussion forums. The first iteration of the pilot, which occurred during Semester 1 (February-July 2019), involved six staff from four broad discipline areas. The second iteration in Semester 2 (July – November 2019) involved a further seven staff from four broad discipline areas. The initial group meeting provided an overview of LA and reports available in the LMS, and of the study. In the first individual meeting, each participant: determined a question they would like to investigate; and discussed why that was important; how they would measure success; data that would help investigate and actions they would take as a result of the investigation. In subsequent meetings the author checked on their progress and provided support and training.

5 RESULTS AND DISCUSSION

This paper reports on the 1st iteration only. Log data showing access to LA reports in the LMS and responses to a feedback survey, completed by five of six participants at the end of the 20 weeks, comprise the available information. The time each participant spent interacting with Analytics reports on the LMS comes from log files for the iteration of their course prior to involvement in this study and for the course during the study. Five of six participants increased their interactions, with Leslie more than doubling their time interacting with the reports. Kendall, whose time spent on LA reports decreased, noted the most frustrating aspect of participation as *“Forgetting things. Because too many other things to do. Consequently, not spending enough time to continue using the tool.”*

Feedback was generally positive with increased awareness and knowledge of how to use LA featuring in comments from all respondents. All respondents identified ways in which they had, and would continue to, use LA to inform teaching practice and course design. Time and ease of access to data comprised barriers. Sample comments from each respondent are included here to show the diversity of responses and the depth to which participants have thought about LA and the process of being involved in this intervention. In response to question on the most rewarding aspect of involvement, Jackie noted *“The structuring of the process ensured connection to undertaking the actual analytics.* Jordan noted the successful use of analytics was *“If the analytics support you to approach considerations of course interactions in a more informed light!”*, whilst Finlay noted the most helpful aspect of involvement was *triggering me to consider different ways of potentially interpreting the analytics.* In response to the most beneficial aspects of involvement, Kendall noted *“Using the USQ Analytics to obtain information that can help me in improving my teaching practices. It is useful to find out where my future emphasis should be placed in terms of teaching the course”*, whilst Keegan responded *“a better understanding of what students are doing”*. Whilst the barriers of time to engage deeply, and difficulty of accessing data, were mentioned by all five respondents, lack of knowledge and training were no longer considered as barriers, suggesting the intervention had

been successful in meeting these needs. Further investigation will consider what support and training will be needed by Learning Designers and Academic Developers to be able to adopt this style of approach more widely. Embedding this level of support into everyday learning and teaching support and capacity-building will enable a more widespread roll-out and uptake across an institution.

This study is limited as there were a small number of participants in one institution and LA are confined to data from the LMS. However, all students in the institution must access the LMS on a regular basis and all assignment submission is conducted through this system. Throughout the pilot, conversations directed participants to further sources of information they could consider, although it was acknowledged that access could be difficult. The depth of discussions and support provided through this study have proven beneficial. Detailed thematic analysis of meeting transcripts will provide more insights on strategies to promote and embed this type of LA initiatives.

6 IMPLICATIONS FOR FIELD

This novel approach to LA adoption has shown that the BCW is an effective approach that institutions can use to develop their own adoption strategies allowing them to strategically provide support and infrastructure to best meet their contextual needs and stage of LA adoption. Whilst staff in each context will have a unique combination of capabilities and motivations and the institution will offer a different set of opportunities, following the processes will enable development of an intervention that includes an appropriate combination of intervention functions and BCT that is most likely to succeed. Sustainability and more large-scale adoption can be achieved by training and involving staff in educational support roles and adopting a distributed leadership approach, calling on early adopters to become champions. Gaining insights into the barriers that still exist, and particularly the types and formats of reports and data that staff find useful is important information for institutions and LMS developers. A growing list of these requests is being created through this study and it is hoped that sharing this will result in improvements, for the participants in this study, the wider population in this institution and more broadly.

REFERENCES

- Bichsel, J. (2012). Analytics in higher education: Benefits, barriers, progress, and recommendations: EDUCAUSE Center for Applied Research
- Gunn, C., J., McDonald, Donald, C., Milne, J. & Blumenstein, M. (2017). *Building an evidence base for teaching and learning design using learning analytics*. Ako Aotearoa.
- Jones, H. (2019). Barriers, enablers, and motivations for staff adoption of learning analytics: Insights for professional learning opportunities from an Australian university. In Y. W. Chew, K. M. Chan, and A. Alphonso (Eds.), *Personalised Learning. Diverse Goals. One Heart. ASCILITE 2019 Singapore* (pp. 149-158).
- Klein, C., Lester, J., Rangwala, H. & Johnri, A. (2019). Technological barriers and incentives to learning analytics adoption in higher education: insights from users. *Journal of Computing in Higher Education* <https://doi.org/10.1007/s12528-019-09210-5>
- Michie, S., Atkins, L., & West, R. (2014). *The Behaviour Change Wheel: A Guide to Designing Interventions*. Silverback Publishing, Great Britain
- Rehrey, G., Goth, D., Fiorini, S., Hostetter, C. & Shepherd, L. (2018). Implementation of a Student Learning Analytics Fellows Program. *Companion Proceedings 8th International Conference on Learning Analytics and Knowledge (LAK18)*.

Building a data warehouse for multimodal learning analytics research projects

Vlatko Lukarov, Matthias Ehlenz, Ulrik Schroeder

Learning Technologies Research Group (LuFG 9) RWTH Aachen University
{lukarov, ehlenz, schroeder}@informatik.rwth-aachen.de

ABSTRACT: This work provides a practical approach for designing and implementing a data warehouse (learning records store) which supports interconnected research projects from the research domain of learning analytics. The research projects come from the learning technologies domain and each one of them has learning analytics as an integral component. The aim of this practical work is to provide a common basis and sustainable infrastructure that manages learning experiences in a structured and safe environment which still provides research freedom and broad accessibility for exploring and analyzing learning experiences.

Keywords: LRS; Multimodal learning events, data warehouse

1 INTRODUCTION

Building technical research infrastructure which can support several research projects is a multifaceted and complex task which usually requires a combination of technical expertise, scientific knowledge and research experience. In many research projects in the field of Learning Analytics (LA) the researcher usually has the role “jack of all trades” who takes care of the technical setup, starting from the learners’ data generation, data preparation and management, data analysis and visualization, all the way up to the conceptualization of experiments, their evaluation and interpretation of the experimentation research. In our case, we have several research projects which research different topics, experiment and test different hypotheses within the context of learning technologies. The common part of all these projects are the LA components which in many cases use the same raw learners’ generated data and events from the used learning technology tools within the learning scenarios and environments. The basic idea of this practitioner’s paper is to share practical experiences by building a common technical infrastructure for the learners’ data, so that each research project (or researcher) does not invest effort and knowledge in generating separate data management strategies and developing different technical infrastructures to implement these data strategies. Apart from the few data challenges and workshops on the LAK conferences, there are very few practical experiences and reports about which factors one has to consider when implementing a technical infrastructure for managing learning experiences for multiple LA projects. The paper is structured as follows: section two outlines our research scenarios and some basics about learning records store, section three outlines the criteria to consider when building a data warehouse, and section four concludes this report with our results and findings.

2 RESEARCH SCENARIOS

The researchers in our group are building and experimenting with innovative learning technologies that shape new digital classrooms and learning and teaching experiences. We have built an e-learning

lab equipped with several virtual reality headsets, several eye-tracking hardware tools, several large multi-touch tabletop displays for collaborative learning games, wide angle cameras and motion sensors. Additionally, we use existing learning technologies such as learning management systems, in-house developed web-based learning tools, and use an extended set of android tablets and mobile phones available for experimentation and learning. The learning activities within the lab generate digital traces which have to be transformed into versatile learning activities for later analysis. We have identified that we need to capture, process and aggregate multiple signal sources to produce events that describe the learning activities and user interactions within the learning processes. In our learning and research scenarios, the traditional log-file learning data collected by an online systems (LMS) are combined with learning events and artifacts which might include physical presence and interactions, multi-touch gestures, gaze, focus, speech, and video recordings. This combination of learner traces from different data sources into a single analysis is the main objective of a subfield called Multimodal learning analytics. The characteristics and properties of these learning contexts cannot be described by a single source of data traces, but a combination of several modes and sources are vital in understanding these particular learning processes (Ochoa, 2017). Similarly, when analyzing the definition and scenarios from Ochoa (2017), and Worsley and Blikstein (Worsley & Blikstein, 2015), our research scenarios and research goals fall under the scope of multimodal learning analytics research.

2.1 Learning records store basics

The most critical part of building our own learning record store was to make an informed decision to choose the most suitable learning records store for our research scenarios. For this purpose we developed a set of criteria (Table 1), with which we analyzed 14 LRS Implementations. The criteria were based on existing technical experiences, existing knowledge and expertise within the team, cost-benefit factors, and long-term viability of the data warehouse. The data warehouse uses the xAPI data standard as an underlying model for storing data because it standardizes the way the learning data and statements are saved from the different (software and physical) sources. All of the different devices, cameras, and sensors, the AR/VR tech, the different table-top interactions, and the LMS generate log events in a completely different structure, frequency, and modality, several xAPI translators and extensions have to be developed and implemented to capture and store the learning events. Additionally, we are developing an appropriate set of verbs and activities which do not exist within the research and practitioners' community for the purpose of building well-defined xAPI statements. This way the modalities are captured in a standardized way which helps in retaining the context and detailed granularity of the learning activities. Once the learning events are generated and transformed into statements, they are stored in the data warehouse via available REST API. The combination and correlation of this multi-modal learner data is aggregated and properly stored within one central Learning Record Store (LRS) and the data should retain its semantic value, and ensure that the data is available for analysis and interpretation while conforming to current data privacy regulations within the EU.

3 RESULTS AND FINDINGS

The first phase of research was an extensive web search of LRS using various search engines. Main criterion on that point was that it had to be mature enough to be at least in beta state and more than a hypothetical construct or hobbyist leisure project. This search was carried out in the first two weeks

of May 2019 and yielded 14 results. We considered the following LRS implementations: Learning Locker, Grassblade, Watershed Essentials, Rustici LRS, OpenLRW (former OpenLRS), Trax LRS, Veracity Learning, IxHive, J18Cloud, Yet xAPI LRS, Riptide Storepoints, Mzinga LRS, edTotal, and ADL LRS.

Table 1. Criteria for choosing a learning records store.

Criteria	Description
License	We need a license in the spirit of the GPL because it should not incur additional costs. For our scenarios CC, MIT, BSD are more restrictive.
Type of service	Self-hosted is the best option for our scenarios because it provides complete hardware and software control over the infrastructure and the data
Data standard	The most suitable solution for us is xAPI as an open standard with suitable extensions
Scalability	We have (for now) four scenarios and in each scenario we have at least four different modalities for the data sources and when all of the generated data is combined together, a lot of physical data is generated. The warehouse has to be easily scalable to handle these amounts of data. (Horizontal scalability is a must to make it easier to add new physical servers/data storage on the fly).
Performance	The warehouse must be able to analyze and support concurrent use, analysis, and delivering results over network since we have several concurrent projects running.
Community	Our goal is to provide back to our research peers and community and share our work. As an added bonus: It is also easier to debug and fix it if something goes wrong.
Technology stack	Well known technologies are easier to develop, upgrade, and maintain.
Data privacy	The LRS should already have implemented and integrated mechanisms which make it data privacy conformant in Germany and the EU. This means built-in features that follow the rules from the state, federal data privacy laws, and the EU's GDPR.
Authentication	Appropriate components for authenticating users, devices, and systems for safe and reliable storage of learning events is a hard requirement.
Extensibility	Important for future integration of new projects and new learning events.
Up-to-date	Important factor that shows how well the infrastructure is updated/used within the community of researchers and practitioners.

3.1 LRS Criteria Application

Our first criterion, open-source software already eliminated a few. Due to serious privacy concerns (GDPR) only self-hosted solutions were considered for our own implementation. Those criteria eliminated Grassblade, Watershed Essentials, J18Cloud, Riptide Storepoints, Rustici LRS, Yet xAPI LRS and edTotal. To ensure long-term support and interoperability the next elimination criteria have been full xAPI standard conformance as well as an open, supportive community and active maintenance. The full xAPI standard conformance is taken as-is from the declaration of the developers, the latter two were achieved by researching deployment in known institutions, activity of support-channels and meta-information from community activities on Git, as well as observation of commits to the official source code repositories. Here OpenLRW, TraxLRS and IxHive dropped out of the comparison because we could not identify traction, activity, or active community development engagement. Finally, Veracity and Mzinga did not openly provide sufficient information to be considered as viable solutions. The decision therefore fell on Learning Locker because it was the closest in fulfilling our criteria.

3.2 Deployment strategies

Deployment of Learning Locker is straightforward. We recommend using one of the various Docker images or a virtual machine running Ubuntu LTS with the official installation script to get accustomed with the system and locally discover the system services. Regarding final deployment our strategy has been a Debian Buster VM to integrate seamlessly into our existing infrastructure. The installation script does not run on Debian out-of-the-box, but custom installation has been the preferred way

anyways to connect the instance to a (local) scalable mongodb cluster on our Docker host machine. Setting up this environment was pretty straightforward as well, beyond basic installation there was just some cross-origin-resource-sharing (CORS) configuration in the VMs nginx necessary. Furthermore, we set up a Virtual Machine (VM) on a laptop configured as an access point and learning app server (first as a proof of concept, later to be integrated within our server infrastructure). This setup provides a completely independent local environment for the web-based tools. This provides reliable conditions for experimental setups (including on the go), as well sufficient “upstream” data rates as context-enriched statements. For example in one project including multi-touch learning games applications, the learning events can be as big as 400 Kbytes, and continuous upstream of learning events generates a data stream of 5 MBit/s and above. This type of setup could heavily impact performance in single-threaded JS applications through dangling promises, and needs careful planning whenever intensive learning interactions are recorded from multiple learners on the go. Similarly, when generating learning events from a VR environment can result in many (large) events which have to be transferred in the LRS. VR environments have a refresh rate of 90Hz+, and include user movement, gaze, location, focus, and relative position within the environment. Events generation takes place in real time and whenever an individual learning event is generated (as part of a series of events), all of these attributes have to be included and the events’ stream pushed into the learning store. Hence, collecting LA data in a local setup with a configured LRS setup allows both live analytics on scene as well as later synchronization with our central instance.

4 CONCLUSION

To conclude our lessons learned, we highly recommend to put enough effort in the planning phase when setting up a learning analytics infrastructure. Identify your use cases, derive your criteria from there. Order your priorities (not everyone might only consider open source like we do), plan ahead and keep your data backend scalable. It could be beneficial to look into used technology stack and go for something you are used to. Get to know your local constraints, both regarding hardware (storage, bandwidth) and legal (GDPR is not really cloud compatible). Identify special use cases like local, off-scene usage and think about synchronization and be creative about possible solutions. Last but not least plan your deployment strategy. If you can afford a dedicated bare-metal setup you might go for it. Docker is great for fast and easy evaluation of new software and usage in development stage and shows its strength in co-hosting multiple services on a machine. We decided on a dedicated VM for possible migration. Keep in mind that according to specification xAPI-statements might be voided but never deleted. Plan for growth of learning events, and infrastructure and accommodate both of them.

REFERENCES

- Ochoa, X. (2017). Multimodal Learning Analytics. In C. Lang, G. Siemens, A. F. Wise, & D. Gasevic (Eds.), *Handbook of Learning Analytics* (First Edit, pp. 129–141). <https://doi.org/10.18608/hla17>
- Worsley, M., & Blikstein, P. (2015). Using learning analytics to study cognitive disequilibrium in a complex learning environment. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK '15*, 426–427. <https://doi.org/10.1145/2723576.2723659>

Engaging Students as Co-Designers of Learning Analytics

Juan Pablo Sarmiento

New York University
jp.sarmiento@nyu.edu

Fabio Campos

New York University
fabioc@nyu.edu

Alyssa Wise

New York University
alyssa.wise@nyu.edu

ABSTRACT: As Learning Analytics (LA) moves from theory into practice, researchers have called for increased participation of stakeholders in design processes. The implementation of such methods, however, still requires attention and specification. In this report, we share strategies and insights from a co-design process that involved university students in the development of a LA tool. We describe the participatory design workshops and highlight three strategies for engaging students in the co-design of learning analytics tools.

Keywords: participatory design, student-facing dashboards, higher education.

1 INTRODUCTION

While the use of Learning Analytics (LA) in higher education institutions is increasing, not many of these tools are intended for students' direct use (Bodily and Verbert, 2017). When such student-facing tools are created, research shows that there can be misalignment between designers' intentions, students' perceptions, and institutional limitations, with learners often distrusting the tools as a result (de Quincey, Briggs, Kyriacou, & Waller, 2019). This may be because the majority of such tools are developed without the direct involvement of students (Bodily and Verbert, 2017). In response, researchers have called for increased participation of stakeholders in the design processes of these tools (Ahn, Campos, Hays & DiGiacomo, 2019; Buckingham Shum, Ferguson & Martinez-Maldonado, 2019).

At New York University we have recently embarked on such a participatory design effort, developing a learning analytics tool *with* and *for* students. This report describes a series of co-design workshops held as part of the participatory process, highlighting key strategies for engaging students in co-designing LA tools for their own tools.

2 INSTITUTIONAL BACKGROUND

New York University prides itself on being an innovator in higher education. With over 50,000 students in both undergraduate and graduate programs, it has begun developing multiple LA initiatives to support its diverse and growing student body. This includes the creation of a LA faculty service as part of the university-wide instructional technology offerings, and the creation of an organization devoted to LA research, NYU-LEARN research network. These two entities have been collaborating on a variety of projects, and in November 2017 embarked on a joint project for the research and design of a student-facing learning analytics tool, with the intent of eventual roll out to the entire student body.

3 PROCESS FRAMING AND PARTICIPATORY DESIGN WORKSHOPS

Our team, composed of members from both the LA service in IT and the research network, was tasked with leading the process of developing a student-facing LA tool which would address challenges for student learning at our university. The team formed a steering committee, composed of stakeholders

from NYU-LEARN, Faculty of Arts and Science and student advisors, which was responsible for setting the major boundaries of the project and participant sampling strategies. The committee decided to target students with diverse backgrounds, focusing on first-generation participants as “extreme users,” whose needs are often underserved when designing for an “average” user (Pullin & Newell, 2007).

Our design process utilized elements of Human-Centered Design, bringing stakeholders in not only as sources of information but as co-designers. The aim was to understand their experiences, needs and points of view, while taking into account learning challenges previously identified by the university. Prior to the co-design sessions, we conducted in-depth interviews with 13 students, three faculty and six advisors to surface the specific needs and challenges faced by students which LA could potentially address. These were used to develop user profiles (“personas”) that outlined salient problems for students’ learning experiences. Students were then invited to take part in co-design sessions to develop LA solutions that tackled these user personas’ challenges.

Participants were recruited through both an open call and individual emails which targeted a list of 106 first-generation students identified by advisors as potentially interested in design. Recruitment was challenging due to time constraints of students at the end of term. Although more than 20 signalled interest, only 10 were able to participate. Most participants came back for multiple design workshops and many indicated explicitly that they were happy with the experience. Students who participated said they did so because of an interest in design and/or big data.

We conducted three 5-hour design workshops, in which a total of 10 freshman, junior, and graduate students (between 4 and 7 in each session) learned about design and LA, and worked alongside researchers and a professional designer. The workshops consisted of: (1) Ice breakers; (2) Short lecture on Design Thinking and Human-Centered Design; (3) Collaborative fleshing out of student personas through empathy mapping; (4) Collaborative mapping of data types and representations relevant to the personas; (5) Individual ideation of solutions through brain-writing; (6) Selection of promising solutions; (7) Collaborative ideation and paper prototyping; (8) Iteration and development of new prototypes (Fig. 1). Facilitators and a designer worked side by side with participants in these activities, iterating between sketching designs, sharing them, providing feedback and redesigning.

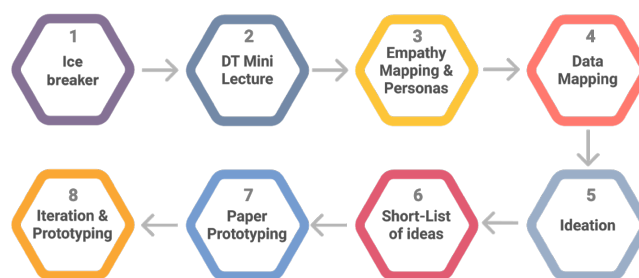


Figure 1: Structure of co-design workshops

To respond to varying levels of data literacy among participants, part 4 of the session was devoted to mapping, alongside students, the types of data the university could collect that might be relevant to the challenges participants identified. Similarly, as a starting point for an ideation brainwriting session (where participants wrote long lists of ideas for solutions), students were presented with a set of cards (Fig. 2), which had (a) inferences that could potentially be made based on the available data (for example “which readings have you opened?”), and (b) action phrases which suggested potential uses for such information (such as “compare”; “help you understand”). Participants randomly took a couple

of the cards from each group and put them together, to stimulate lateral thinking and simultaneously ground their ideas in the available data. They would then share their favorite ideas, and iterate on these ideas in teams.

The resulting designs ranged broadly, from a tool to aid students in finding “study buddies” to systems where users who took a given course could share information with new students. Students’ solutions were often similar to those described in the literature (like tools for monitoring course progress or using peer activity as a motivator) and sometimes pushed in directions new to the field (like leveraging social networks and emotion as central elements of design).

Participants also brought their identities and interests into the discussion. For instance, one student advocated for platforms which took into account variables such as wellbeing and mental health, after having themselves been diagnosed with depression. Some ideas were mixed with others in playful ways, often using metaphors such as a “hive” in which students visualized competing deadlines and responsibilities from multiple courses, merged with a timeline, and a system for monitoring peer resource activity (Fig. 2). Students also identified qualitative input from other students (such as advice, tips or warnings) as a valuable source of information for how to successfully navigate personal and academic challenges. While LA typically concerns itself with the presentation of data for insights, this suggests that systems that pair analysis with the communication of know-how information through a human network can have special value for students.

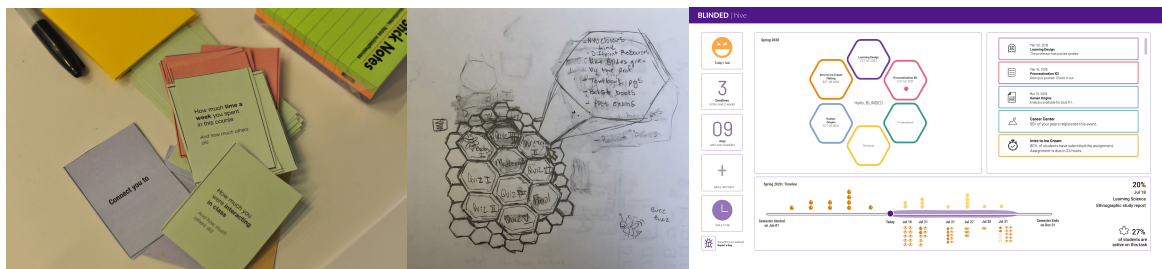


Figure 2: Ideation cards, student sketches and first prototype of the Hive interface.

4 KEY STRATEGIES FOR ENGAGING STUDENTS

Below we highlight three key strategies we found useful in the participatory design workshops:

Explicitly Address Power Dynamics. Power imbalances, such as those between students and researchers or designers, can hinder participatory processes (Dollinger, 2018). Developing trust between participants can be a way to overcome this problem. Throughout the workshops, facilitators reminded participants that it was safe to share opinions and to be provocative, and encouraged students to challenge facilitators’ views and assumptions. The strategy seemed to succeed, in that participants scrutinized the very objective of the session. They pointed out that questions of happiness and meaning may be more relevant to students than academics, inciting a discussion about the role of a university in students’ lives.

Keep the Problem Space Flexible. Traditional design thinking advocates for a clear problem framing, but when working with students we saw value in allowing the room to go back to problematizing the question, and refrained from limiting students’ ideation even while hunting for solutions. When students pointed to unexpected problem spaces, such as happiness vs. academics, or the role of family and motivation, these were honored. We believe that this played a role in them framing and

integrating solutions in novel ways, such as tools which used interactions between student's academic and emotional data to provide suggestions.

Use Vulnerability to Develop Psychological Safety. Psychological safety can be an important precondition to creativity (Hunter, Bedell & Mumford, 2007), though one hard to achieve with strangers. As facilitators, we used sharing vulnerability as a means to create a safe space, through ice breaker exercises where participants worked together, engaged in physical contact and were silly in front of each other. Likewise, participants were encouraged to share personal experiences and stories. Success of the strategy is evident in video recordings of the workshops, which show participants shifting from being hesitant to contributing to laughing and enjoying working together. The environment is likely to have played a role in students being increasingly comfortable with sharing unconventional paths to a solution, such as a tool where achievement was monitored through the "health" of a virtual pet (like a Tamagotchi). Several of these more unconventional solutions emphasized affective and playful design, an area that has yet not been central in LA tool creation.

5 CONCLUSION AND NEXT STEPS

As some in the field have argued, having students actively participate in the process of co-creating LA tools can have a positive impact on ownership, understanding and use (de Quincey et. al., 2019; Dollinger, 2018; Buckingham Shum et al., 2019). Our experience with participatory design suggests that addressing issues of power imbalance and developing an environment conducive to creative thinking has the potential to uncover innovative designs with the potential to improve students' experience and learning. At the time of this writing, we are preparing to test this hypothesis through user experience trials of prototypes developed from these workshops, in order to understand if the process indeed produces student-facing LA tools whose purpose is valued by students and which are seen as aligning with their authentic needs.

REFERENCES

- Ahn, J., Campos, F., Hays, M., & DiGiacomo, D. (2019). Designing in Context: Reaching beyond Usability in Learning Analytics Dashboard Design. *Journal of Learning Analytics*, 6(2), 70-85.
- Bodily, R., & Verbert, K. (2017). Trends and issues in student-facing learning analytics reporting systems research. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 309-318).
- Buckingham Shum, S., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-Centred Learning Analytics. *Journal of Learning Analytics*, 6(2), 1-9.
- Dollinger, M. (2018). Technology for the scalability of co-creation with students. *Open Oceans: Learning without borders*, 346-350.
- Hunter, S. T., Bedell, K. E., & Mumford, M. D. (2007). Climate for creativity: A quantitative review. *Creativity research journal*, 19(1), 69-90.
- Pullin, G., & Newell, A. (2007, July). Focussing on extra-ordinary users. In *Proceedings International Conference on Universal Access in Human-Computer Interaction* (pp. 253-262).
- de Quincey, E., Briggs, C., Kyriacou, T., & Waller, R. (2019, March). Student Centred Design of a Learning Analytics System. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 353-362).

Considerations for amending a whole-institution early-alert system

Rebecca Siddle

Nottingham Trent University
rebecca.siddle@ntu.ac.uk

Ed Foster

Nottingham Trent University
ed.foster@ntu.ac.uk

ABSTRACT: Nottingham Trent University (NTU) has had a whole-institution learning analytics-based alerting system since 2014-15. These alerts were based on 14 consecutive days of non-engagement during term time, and were designed to be marked indicators of risk. In September 2018, a new version of the learning analytics platform was released which allowed more flexible alerting options. This paper outlines the design-, data- and ethics-driven decision making required to agree new institutional alerting parameters. These decisions were based on a combination of the institutional context and the calculated non-progression rates of students generating hypothetical alerts of different lengths during 2017-18.

Keywords: Learning analytics, Early-warning, student success

1 INTRODUCTION

A learning analytics system, termed internally the 'NTU Student Dashboard' (Dashboard), was implemented at the institution as a means to improve student outcomes. The system focuses on student 'engagement' and includes an early-warning alerting system. The alerts are designed to inform person tutors if their tutees have disengaged from the University for a sustained period during term time, so the tutor can contact the student and offer support. The resources used as proxies for engagement are as follows: Virtual Learning Environment (VLE) logins, VLE Learning Room logins, Dropbox submissions, library loans, e-resource usage, card swipes and attendance. Further information on the implementation of learning analytics at NTU, research around previous years' no-engagement alerts, and relevant literature references can be found elsewhere (Foster & Edwards 2018, Foster & Siddle, 2019).

In September 2018, the commercial provider released a new version of their platform that had more flexible alerting options. This allowed institutions to amend the following alert parameters: the level of engagement, the time period, and the group of students. The seemingly simple task of ascertaining what the alerting parameters should be for 2018/19 took the team back to the fundamental principles and considerations of applied learning analytics.

The work has been conducted as part of an Erasmus+ funded project entitled 'Onwards from Learning analytics' (OfLA), and continues work from two previous Erasmus+ funded projects: ABLE

and STELA. Readers interested in staff and student views on the Dashboard and related initiatives should consult the relevant resources on the project websites.¹

1.1 Analysing the current situation

The original design-decisions reflect the institutional vision of alerting and the desire to ensure alerts are easy for tutors to comprehend and subsequently communicate to students. Alerts were designed to indicate students at high risk of non-progression, and to act as an immediate call to action for staff. This focus on high risk was felt to be appropriate because the alerts are sent in addition to a broad range of support available to serve the wider student population. The alerts were 'no' engagement as this requires no value judgement about the level of activity. The alerting period was set to 14 days. This timeframe was selected because early analysis showed that students with 14 days of non-engagement were significantly less likely to progress than their peers, and also because it was felt to represent an appropriately sustained period of non-engagement given the average length of each term (approximately 10 weeks) to merit contacting students. Cognisant of the institutional scale of the set-up, certain students were omitted based factors such as their enrolment status and mode of study. Alerts are sent based on term-time engagement only.

1.2 Considering the potential advantages/disadvantages of change

Having the alert timeframe standardised across all year groups has the advantage of being simple to comprehend for tutors, and provides consistency for students. However, our statistics show that second and final year students are more likely to progress than their first year counterparts. This variation in outcomes led us to consider tailoring the alerting by year group. From a risk of progression perspective it may be appropriate, indeed preferable to do so, but it requires a ethical decision about whether treating students from these year groups differently is fair.

Shortening the timeframe of the alerts means they can be sent out earlier. This could be the difference between reaching a student whilst they are in the process of disengaging and once they've fully disengaged. It also maximises the opportunity for intervention during term time. However, shortening the timeframe of the alert also increases the number of alerts that can be generated about an individual student in a given period. These alerts may serve as useful reminders to tutors, but repeated alerting may be seen as spam, and hence may be more likely to be ignored. This would undermine the purpose of alerting.

2 METHODOLOGY

Hypothetical alerts were calculated for students based on their daily engagement activity during 2017/18. Alerts were calculated on a rolling-basis so could be generated any day of the week upon the student reaching the required number of consecutive days of non-engagement. The data reported is based on the group of students who generated one or more alert. Initially, 7- and 10-day alerts were studied as the potential alternatives to 14-day alerts. Following initial analysis, 21-day alerts were also studied for final year students.

¹ <https://oflaproject.eu/>, <http://www.ableproject.eu/>, <https://stela-project.org/>

The engagement data used in this report was produced by the commercial provider's tool (Solutionpath's StREAM, Version 4). Student cohort details (course, year group), enrolment and progression data was downloaded from the University's data mart, Cognos. A student is not considered to have progressed if they transferred to the same year of a different course, withdrew or failed academically.

3 RESULTS AND DISCUSSION

The success of the alerting timeframe was assessed based upon the balance between the accuracy of the alerts as predictors of non-progression, and the proportion of the total non-progressing student population that the alerts would identify.

The data below focuses on the progression of students who would have generated alerts if the alerting timeframe was set to 7-, 10-, 14- or 21- consecutive days of non-engagement, see Table 1. Combining the total number of students with the percentage progression, it is clear that shorter timeframes for alerts identify a greater number of non-progressing students, but also a greater number of progressing students. In 2017/18, 82% of first year students progressed compared with 90% of second year students and 95% of final year students. This means that all alert timeframes tested identified a group of students with lower rates of progression than the average for their year group. Furthermore, all alerting timeframes are shown to be less severe indicators of risk of non-progression for final year students than second or first year students.

Table 1: Progression of first, second and final year students who would have generated one or more 7-, 10-, 14- and 21-day no-engagement alerts during the 2017/18 academic year.

Year group	Timeframe of alert	Total students	Percentage of students (%)	
			Not Progressing	Progressing
First year	7 days	1,625	42	58
	10 days	870	60	40
	14 days	448	78	22
Second year	7 days	1,795	23	77
	10 days	1,001	34	66
	14 days	546	55	45
Final year	7 days	1,331	15	85
	10 days	510	30	70
	14 days	367	36	64
	21 days	192	51	49

The alerts are designed to prompt an offer of support. Although this may seem to be an inherently positive action, there can be negative consequences if students are wrongly identified. Generating false alerts about students has the potential to put unnecessary strain upon staff resource, to take staff focus from students most in need of support, to reduce staff trust in the alerting system, and to wrongly stigmatise, demotivate or worry students who are misidentified (Foster & Siddle, 2019).

What is felt to be a reasonable risk of student non-progression to act upon is likely to vary by person, and to be influenced by the resource available. When considering the most appropriate alerting timeframes to set at an institutional level, the team reviewed the role of the alerts and the

mechanism through which they were due to have impact. It was agreed that the alerts should remain as indicators of high risk and that the parameters for determining this should focus on the human aspect of the alerts. In the context of everyday life, 50% chance could be considered as a psychological tipping point: less than this and a prediction is less accurate than guessing heads/tails on a coin toss. The design decision was taken to build the alerting around a greater than 50% chance of non-progression. The data suggested that if a single timeframe of alerts was to be selected for all year groups 14-day alerts should remain the standard length, as 57% of all students who would have generated one or more 14-day alert failed to progress. Varying the length of alerts would allow the percentage progression to be above 50% for all year groups, by setting first year alerts to 10 days, second year alerts to 14 days and final year students to 21 days (non-progression rates of 60%, 55%, and 51% respectively). Overall, varying the length of alerts would result in the same proportion of non-progressing students (57%) as standardised 14-day alerts, whilst identifying 18% more non-progressing students (1,361 and 1,608 students respectively).

Senior decision makers were provided with the data and the recommendation to vary the alerts by year group if comfortable with treating the year groups differently. Shortening the timeframe for first year students was agreed, due to increased risk of non-progression and the expectation that these students would be less familiar with the University's support systems, would have less well established local support networks, and would have developed fewer strategies for succeeding in HE. However, despite the lower risk, it was felt that extending the timeframe for final year students would not be fulfilling our duty of care responsibility, as three weeks was too long to leave a student who may be in need of support. As a result, the alerting parameters chosen for 2019/20 were 10 days for first years and 14 days for second and final years. This is a clear example of where data and ethical arguments have been combined to agree the outcomes.

4 CONCLUSIONS

This paper outlines the considerations for defining new, institution-wide alerting parameters at NTU. The decision to vary the timeframe of alerting by year group represents an additional level of sophistication, although we recognize the scale of further work required to be able to address the criticisms of early warning systems in the literature (Cano & Leonard II, 2019). We anticipate further sophistication will be possible as additional institutional data sources become systematised for learning analytics.

REFERENCES

- Foster, E., & Edwards, R. (2018, March). *Implementation of an institution-wide learning analytics dashboard: a case study*. Paper presented at Learning Analytics Knowledge 2018. <http://bit.ly/lak18-companion-proceedings>
- Foster, E., & Siddle, R. (2019). The effectiveness of learning analytics for identifying at-risk students in higher education institutions. *Assessment & Evaluation in Higher Education*. Advance online publication. <https://doi.org/10.1080/02602938.2019.1682118>
- Cano, A., & Leonard II, J. (2019). Interpretable Multi-view Early Warning System adapted to Underrepresented Student Populations. *IEEE Transactions on Learning Technologies*, 12(2), 198-211. DOI:10.1109/TLT.2019.2911079.

Why clicks are not enough: designing a Learning Analytics service for the Estonian Open Educational Resources ecosystem

Kairit Tammets, Tobias Ley, Mart Laanpere and Manisha Khulbe

Tallinn University
kairit.tammets@tlu.ee

ABSTRACT: Open Educational Resources (OERs) have gained ground in the educational landscape, and are being increasingly used by teachers in a flexible way to support classroom learning. However, the pedagogical practices that make use of OERs are not always student-centred. This article summarises the validation of a LA service in the context of Estonian national-level deployment of an OER ecosystem, and analyses its limitations by evaluating the designs of the classroom practices using LA data. The results of this study contribute towards the generalisation and improvement of LA services that are integrated into an open ecosystem of OERs.

Keywords: Open Educational Resources, Learning Analytics, Learning Design

1 INTRODUCTION

Deployment of Open Educational Resources (OERs) in teaching processes is a growing trend. By OERs, we refer to open-access, reusable digital resources used for the purpose of learning, or designed to be used in learning settings (OECD, 2007). Such resources are applicable in different contexts, freely combinable, and openly accessible to be used according to a particular Learning Design (LD). Among other digitally-innovative countries, Estonia has made significant investments in technology-enhanced learning to realize the Digital Turn program in education as formulated in the National Strategy for Lifelong Learning 2014-2020.

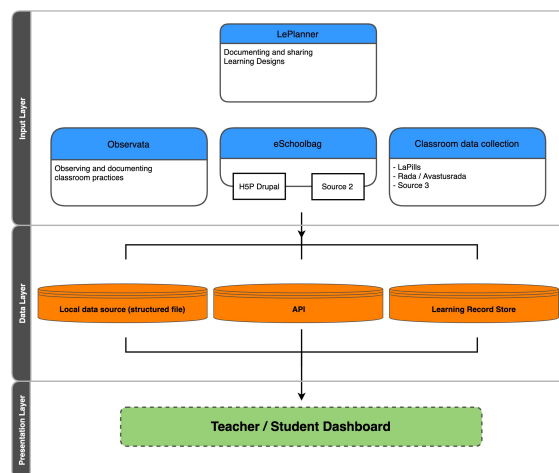
However, some challenges need to be addressed. First, the uptake of OERs requires support—not only technical but also pedagogical (Fullan, 2013). Therefore, pedagogically meaningful LD should be proposed to accompany OERs to maximize their impact on learning. Secondly, while LD is an important means to understand the context of use, the design does not reveal anything about actual classroom practice, which may be the same as or different than planned. Thirdly, the methodology and the architecture to understand the effectiveness of OERs in classroom practice are needed. It has been proposed by Mangaroska & Giannakos (2017) that without contextual interpretation of the data collected with digital technologies, LA capabilities are limited.

This study analyzes the experience of national-level implementation of OERs to evaluate the accompanying LA service, and propose an infrastructure for explaining classroom practices integrating OERs. Paper reports a first validation of LA service by means of a case study. Research questions were formulated as follows: a) to what extent does LA-service aid researchers in understanding classroom practices involving use of OERs and b) what are the components needed for the improved LA service to effectively analyze these classroom practices.

2 LA FOR OPEN EDUCATIONAL RESOURCES

In 2017-2019, a national-level project was carried out to develop interactive task-based OERs that cover the entire upper-secondary school curriculum in four subject areas: mathematics, natural sciences, social sciences, and arts and music. More than 6,000 interactive online learning resources were developed by ca 100 Estonian teachers in collaboration with university experts in subject didactics and educational technology. To develop new original OERs, interactive learning resource templates created using H5P and running on Drupal Content Management System were chosen. The platform was set up, new H5P templates were developed and the Drupal side was also enhanced with management of Learning Object Metadata and OAI-PMH interface for automatic harvesting of the metadata of H5P learning objects by the national learning resource catalogue e-koolikott.ee.

Figure 1: LA service for OERs to explain classroom practices



The LA service (fig 1) is designed for monitoring of the usage of OERs consisting of different technologies used in our study developed for explaining how OERs were integrated in teaching. The input layer of the system consists of LePlanner, Observata, classroom-level data collection tools and OERs. LePlanner is used by teachers for visualizing and sharing pedagogical scenarios by sequencing learning tasks and linking them with relevant OERs. Observata is an extension to LePlanner that enables documenting the classroom observations based on LePlanner scenarios. Observations are stored as sets of xAPI statements; however, observers can add unstructured notes and photos to the observation timeline (Eradze et al, 2015). The next is the data layer, where the important components is the Learning Record Store (LRS), which collects students' interactions with the OERs. LRS Learning Locker 2.0 was used for storing xAPI statements generated by the H5P templates. In this layer the processing of the data from LRS and classroom level data collection tools used by the teacher in monitoring process happens to aggregate, organize and analyze data for the different users. Finally, the presentation layer to visualize data for the users and developed.

3 CASE STUDY

The LA service was validated as part of the national level project "Digital learning resources for secondary education in math, arts and music and social and natural sciences". The case study was conducted from March– May 2018 when 32 teachers piloted OERs with their students (1,200 students). Teachers designed the lesson with LePlanner enabling us to analyze the intended LD. Next, classroom enactment was conducted, when students interacted with the OERs and log-data

was collected to the LRS. To understand what practices the teachers employed, some of the classes were observed by using observation tool Observata. After each class, teachers used online form to answer: what did they plan to do in the class, which OERs were used, what obstacles faced, and what worked well. Besides analyzing data from the whole pilot group, we also analyzed in more depth data about one history teacher who conducted two classes on two consecutive days with two separate group of students. Data from OERs, sessions from Observata, and self-reflections was extracted from these periods of times (morning of the 16th and 17th of April). Intended designs were analyzed from LePlanner, and also from teachers' self-reflections. Qualitative content-analysis of the self-reflections was carried out to identify the pedagogical practices, challenges and opportunities.

4 RESULTS

16.04 08.15 – 09.00	17.04 11.20 – 12.05
<ul style="list-style-type: none"> - 17 unique users accessed the theoretical material - 4 users completed task 1: <ul style="list-style-type: none"> - Max score 5, average 4,61; average time spent 4,5 min - 4 users interacted with task 3 	<ul style="list-style-type: none"> - 21 unique users accessed the theoretical material - 17 users completed task 1: <ul style="list-style-type: none"> - Max score 5, average 4,25, average time spent 11 min - 17 users interacted with task 3

In order to understand if LA service supports explaining usage of OERs in classroom, we followed several phases and integrated different data-layers to our study. In our study, students used the materials through national level repository and did not log into the system. As a result, data was anonymous and personal level interactions were not identified, but data was extracted at the classroom level. As the piloting classes were previously agreed for observations, we knew when the classes took place and based on the timestamp and we analyzed the xAPI statements further (Fig 2).

Fig 2: Analysis of the two sessions based on xAPI statements

Based on the analysis of these two sessions, taught by the same teacher in two parallel classes, it was not clear why the second group of students (17.04) spent more time on completing the task 1 as compared with the 1st group (16.04). Also, we did not understand why only 4 students from the first group interacted with the tasks, whereas in the second group (17.04) almost all the students interacted with the content. To understand the intended design of the lesson, LDs from LePlanner and teacher's self-reflections were analysed (16.04 and 17.04). The teacher planned to implement the same LD with the two parallel 10th grade class: three tasks, one about theoretical material and other two tasks involved focusing on collaboration. Data showed that some activities were group activities (explaining why not all the students interacted with the content). However, we still did not know what happened in the classroom – was the LD implemented as planned, and if not – why. Observations helped to understand that the teacher had to support students quite a lot technically (a lot of 'attempt' verbs in one time period), which had an impact to LD. This is the information we could not extract from the log-data. The teachers' self-reports were the final layer to understand their perspective of the classroom activities. Going back to the example, where we saw that the same teacher implemented the same LD with two different classes, but results of the logs data were different. The teacher's reflection indicated that due to technical issues, she was not successful in

implementing the design (Teacher: *“Some of the students did not manage to interact with the digital resources tasks 2 and 3, materials just did not open”*). This explains why one lesson had a smaller number of unique interactions than the other lesson. Students’ technical issues with the technologies is something we could not detect from any other data source, but these play an important role in implementing LD. We also learnt from that students’ time spent on task and the results were different in two classes – one class worked longer with the materials and scored less (17.04) than the first group (16.04). After the class, the teacher reflected: *“all the planned learning activities took longer with other group of students, because it is bilingual group of learners whose learning pace is much slower.”* This is important information we need to validate the LDs, because with the multi-lingual students, teachers have to design the activities slightly more flexibly.

5 CONCLUSION AND FUTURE WORK

In our study we evaluated the LA service for the Estonian OER ecosystem and learnt that clicks from OER are not enough to explain the classroom practices. Without contextual information regarding how the materials are embedded in the pedagogical practices to activate students, it is complicated to evaluate the efficiency of the OERs. Study demonstrated how better understanding is needed on how learning happens while students are interacting with OERs, therefore we need to design and evaluate the practices with multiple data resources. Such data could be collected through observations, self-report instruments, process-oriented data collection. The proposed LA service will be evaluated in the next phase during the training program for the teachers who co-design novel practices integrating OERs. Also, tool is proposed for researchers, policy-bodies, publishers, and authors to understand the practices of teaching and learning using OERs and their impact.

ACKNOWLEDGMENTS

This research has received funding from the European Union’s Horizon 2020 research and innovation programme under grant no. 669074 (CEITER, <http://ceiter.tlu.ee>)

REFERENCES

- Eradze, M., Rodríguez-Triana, M.J., Laanpere, M. (2017). Semantically Annotated Lesson Observation Data in Learning Analytics Datasets: a Reference Model. *IXD&A Interaction Design & Architecture(s)* 33, 75–91.
- Estonian Ministry of Education and Research. 2014. The Estonian Lifelong Learning Strategy 2020.
- Fullan, M. (2013). *Stratosphere: Integrating technology, pedagogy, and change knowledge*. Pearson, Don Mills, Canada.
- Mor Y., Ferguson, R., & Wasson, B. (2015). Editorial: Learning design, teacher inquiry into student learning and learning analytics: A call for action. *British Journal of Educational Technology* 46, 221– 229.
- OECD. (2007). Digital learning resources as systemic innovation. Project outline and definitions.
- Wragg, T. (2013). *An Introduction to Classroom Observation* (Classic Edition). Routledge.

Open Learning Analytics Indicator Repository

Atezaz Ahmad, Jan Schneider, Hendrik Drachsler

DIPF | Leibniz Institute for Research and Information in Education

ahmad@dipf.de, schneider.Jan@dipf.de, drachsler@dipf.de

ABSTRACT: In this study, we proposed a system of an overview of the indicators in the field of learning analytics over the past decade. The system is based on our literature review from a total of 123 scientific publications, where we fetched 158 indicators, metrics, learning activities, and learning events. We present a system that will provide indicators based on learning activities and learning events. The aim is to help the stakeholders in the application of learning analytics.

Keywords: Indicators, Metrics, Activities, Learning Analytics, Learning Design, Events

1 INTRODUCTION

A challenge that LA faces these days is the “design-benefit” dilemma. Currently, we can track the learning behaviors for cognitive, metacognitive, and even psychomotor learning tasks. However, best practices for aligning the collected data with learning design, and, therefore, obtaining meaningful outcomes are still scarce. The design of most learning environments nowadays does not exploit the benefits and added insights that LA can bring. Currently, it is possible to track a lot of data about learners and their interactions with learning environments; however, it is not clear how to infer relevant indicators out of these collected data (Ferguson, 2012). Moreover, it remains unclear how to present back these indicators to learners and teachers in an intelligible format (Chatti, 2014). For example, if someone wants to apply LA for reading comprehension, it is not clear what data needs to be collected and how to use it to provide meaningful indicators about the students' progress in this particular area. Thus, there is a need for a system that acts as an indicator repository, which informs researchers, teachers and course designers about indicators that have already been successfully applied based on the learning design. In this poster paper we present the concept of a LA indicator repository that facilitates the task of designing a meaningful LA system by presenting to the stakeholders of the learning process suggestions of metrics and indicators that have already been successfully applied in the past. To this end, we did a systematic LA literature review, where we gathered the metadata (e.g., events, metrics, indicators, activities, etc.) reported within. We further aligned LA to the field of learning design (LD) to see how it correlates with LA. The results of our proposed system will be based on the LA and LD activities, this system will also provide a way to evaluate the LD based on the indicators and metrics provided.

2 STATE OF THE ART

Few studies took LA indicators into account to establish a connection between LA and LD. Lockyer, 2013, provided an overall general idea and framework of learning analytics for learning design where the teacher is provided with analytics while designing a course. Also, a study by Martin et al., 2016 underlines that LA can provide evaluative feedback to the design of learning activities. Only one study (Biedermann, 2018) considered the same idea as we suggested, but the approach of developing and presenting the outcomes are very different, where the outcomes are represented in

a directed graph. A criticism of LA concerns its data-driven approach, which does not necessarily always align with the pedagogical aspects of education (Jivet, 2017). To address this issue, we developed a reference framework (see Figure 1) that aligns LD and LA with the aim to improve learning. Our reference scheme consists of two explicit research fields that we consider mutually reinforcing: (1) learning design (LD) and (2) learning analytics (LA). The LD part of our reference framework builds on the study of (Mangaroska, 2018) where authors attempted to align LA with LD. It includes a process-based view on learning objectives, learning events, learning activities, resources/support, learning tasks, and activity outcome.

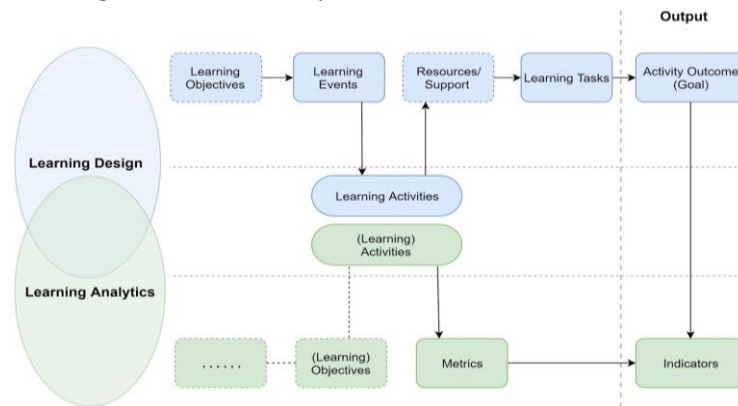


Figure 1: Reference Framework

The second main part of our framework concerns LA. We first extracted the following metadata from LA literature: title, objectives, metrics, indicators, activities, data source, stakeholders, evaluation methods/approaches, dataset size, keywords, country, year, authors, and reference.

By looking at the literature we identified that learning activities play a fundamental role in LD (Craftb, 2012) as well as in LA. Therefore, we use learning activities as a common factor in LA and LD to establish a link and further lead us to meaningful indicators. In our framework these meaningful indicators provide insights into the learning outcomes, hence closing the connection between LA and LD. We argue that these indicators can help learners and teachers to adapt their behavior with the purpose to achieve the desired outcomes, as well as providing course designers feedback about the learning activities.

3 OPEN LA INDUCATOR REPOSITORY

In this section, we present the open LA indicator repository. The open LA indicator repository is a system whose frontend consists of a dashboard. This dashboard will provide an interface that filters out the list of indicators and their metrics that can be used for a particular activity. All our results will be based on the literature that we have conducted previously. Our dashboard will contain learning events, learning activities, indicators, and metrics. Where **Learning Event** is learning or teaching event occurs during a learner's activity or a teacher's activity (Leclercq, 2005). Leclercq and Poumay identified eight learning events: create, explore, practice, imitate, receive, debate, meta-learn, and experiment. **(Learning) activity** is an activity where an action that the learner does in an LMS environment, for example, posting, discussing, uploading, etc. (Duval, 2011). Usually, in the LMS environment, all these activities are captured and stored to analyze and can be used to improve learning (Park, 2015). **Metrics** are the data collected (measurements) about the activity that a learner does in a learning environment (e.g., # of views, # of posts, etc.). **Indicators** are metrics used to create indicators; an indicator is a product of multiple metrics to give a more complete picture on a particular (abstract) learner status, e.g., self-reflection, etc.

Our dashboard will provide a visualization showing how each learning event has multiple learning activities, each learning activity has indicators, and every indicator is generated by multiple metrics. Apart from displaying the indicators that one can use for a particular event and activity, the system aims to provide help in the implementation/execution of the indicator based on the metrics. This open LA indicator repository can help educational practitioners who want to apply LA to their LD. Suppose, if someone wants to apply LA to the learning activities (e.g., exercise, writing, assignment, etc.) using a learning platform, then the open LA indicator repository can suggest this person a list of indicators (e.g., Student performance, Predictive analytics, Writing analytics, etc.) that have already been successfully used.

3.1 Challenges

The development of the open LA indicator repository has a number of challenges. First, how can we define or add more activities, metrics, and indicators on runtime, and merge them into existing activities, metrics, and indicators. Indicators can have different metrics, for example in paper (Coffrin, 2014) performance indicator had metrics A, B, and C. On the other hand in study (Matcha, 2019) performance indicator had metrics D, F, and G. Second, how to present the open LA indicators dashboard in a helpful and effective way for teachers and developers. Third, the platform should provide a dynamic interface and functionality to (semi)automatically retrieve the data from new peer reviewed publications to actualize the and enrich the open LA indicators repository. Last but not least, there should be a mechanism to automatically extract the important data (e.g. activities, metrics, indicators, etc.) from the research papers provided.

REFERENCES

- Leclercq, D., & Poumay, M. (2005). The 8 learning Events Model and its principles (Release 2005-1).
 Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist*.
 Mangaroska, K., & Giannakos, M. N. (2018). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE*
 Martin, F., Ndoeye, A., & Wilkins, P. (2016). Using learning analytics to enhance student learning in online courses based on quality Matters standards. *Educational Technology Systems Journal*
 Park, Y., & Jo, I. H. (2015). Development of the learning analytics dashboard to support students' learning performance. *Journal of Universal Computer Science*, 21(1), 110.
 Duval, E. (2011). Attention please!: learning analytics for visualization and recommendation. *LAK*.
 Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6), 304-317.
 Craftb, B. (2012). Learning Design: reflections upon the current landscape.
 Chatti, M. A., Lukarov, V., Thüs, H., Muslim, A., Yousef, A. M. F., Wahid, U., ... & Schroeder, U. (2014). Learning analytics: Challenges and future research directions. *eleed*, 10(1).
 Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017, September). Awareness is not enough: pitfalls of learning analytics dashboards in the educational practice. In *ECTEL Springer, Cham*.
 Biedermann, D., Schneider, J., & Drachsler, H. (2018, September). The Learning Analytics Indicator Repository. In *ECTEL* (pp. 579-582). Springer, Cham
 Coffrin, C., Corrin, L., de Barba, P., & Kennedy, G. (2014, March). Visualizing patterns of student engagement and performance in MOOCs. In *Proceedings of the LAK*. ACM.
 Matcha, W., Gašević, D., Uzir, N. A. A., Jovanović, J., & Pardo, A. (2019, March). Analytics of Learning Strategies: Associations with Academic Performance and Feedback. In *LAK 19*. ACM.

Examining the Relationship Between Temporality and Social Positions in Collaborative Web Annotation

Rukmini Manasa Avadhanam

University of Minnesota, Twin Cities
avadh005@gmail.com

Bodong Chen

University of Minnesota, Twin Cities
chenbd@umn.edu

ABSTRACT: Web annotation is used to support social reading of online documents. In this study, we used temporal analysis and social network analysis to examine the extent to which the temporal and social dimensions of collaborative web annotation are intertwined within the context of an online class. Using temporal social network analysis, we revealed the evolution of the class's interaction network and each individual student's network measures over one semester. We then correlated each learner's network measures with two temporal measures we constructed to characterize the timing and temporal dispersion of their participation. Results indicated being an early starter and checking back more frequently (i.e., more temporally dispersed) were linked to higher network centralities. Future research is needed to investigate the temporal unfolding of social interaction in learning contexts.

Keywords: Social Annotation, Temporal Analysis, Social Network Analysis, Online Learning

1 INTRODUCTION

The Computer-Supported Collaborative Learning (CSCL) community is interested in promoting learner engagement through digital technologies. Learner outcomes like effective communication, knowledge sharing, and learning motivation have been associated with students using social annotation tools (Mendenhall, 2010; Zurita & Nussbaum, 2004). Although methodological approaches like content analysis and social network analysis were applied to investigating the conceptual and social engagement in social annotation, we argue that the temporal dimension of social annotation in a collaborative learning context needs to be further examined (Chen, Knight, & Wise, 2018). To mitigate the gap, this study integrates temporal analysis with social network analysis (SNA) to probe the connection between learners' temporal participation patterns with their social positions in the interaction network. Within a context of a semester-long graduate seminar where students used a web annotation tool to engage with all course readings, we asked the following research questions:

1. How did each learner's network positions in the class-level interaction network change over the semester? How did their ego-networks evolve over time?
2. To which extent did the learner's timing and temporal dispersion of annotation activities associate with their network positions?

2 METHODOLOGY

2.1 Research Context

The research context was a graduate-level 14-week online course offered at a large public university in the US. Course participants ($n=14$) used Hypothes.is, a web annotation tool, throughout the semester to collaboratively annotate weekly readings and discuss key concepts in the readings.

2.2 Data Sources

Following an IRB approval, a total of 1,890 annotations that were produced by participants were compiled for data analysis. The course syllabus and website were used to contextualize data analysis.

2.3 Data Analysis

2.3.1 Temporal Social Network Analysis

Using the Hypothes.is annotation data, whole-class interaction networks were constructed based on the reply connections. Network visualizations and measures (e.g., *density*) were created using Gephi. To answer the first question, we sliced the interaction network by week and constructed ego networks for each week. To depict a learner's network position in the whole-class network, we computed node-level network measures including in-degree, out-degree, and closeness centralities; to characterize one's local social structure, we computed the *density* of their ego network.

2.3.2 Temporal Measures

To answer the second research question, we first calculated two temporal measures based on the timestamps of learner annotations. For each annotation, we calculated the time difference (Δt) between the time it was created and the week's starting date. Using Δt , we further computed two measures to characterize the distribution of a learner's annotation activities across the semester.

- A. *Temporal Dispersion*: To capture the extent to which a learner's annotations were dispersed across each week, we calculated the standard deviation of Δt of the learner's annotations.
- B. *Timing of participation*: To capture the timing of a learner's participation, we first computed the mean of each learner's time difference (Δt) and then divided learners into two categories – early and late starters – based on the measure. If the mean was below 84 hours (the midpoint of each week), the learner was classified as an early starter; and vice versa.

To gauge the connection between a learner's social position and temporal participation patterns, Spearman's correlation coefficient was used for correlating network measures with temporal dispersion and point-biserial correlation was used to correlate network measures with the dichotomous variable of timing.

3 RESULTS

The whole-class network's density ranged from 0.07 to 0.25 across all weeks ($M=0.13$, $SD=0.09$). For most individual learners, out-degree centralities were consistent across all weeks but in-degree centralities varied within a range of 1 to 12. For instance, looking at the in- and out-degree centralities of learner #4, the learner has constantly sent three responses to their peers each week, but has received responses ranging from one to twelve throughout the course period.

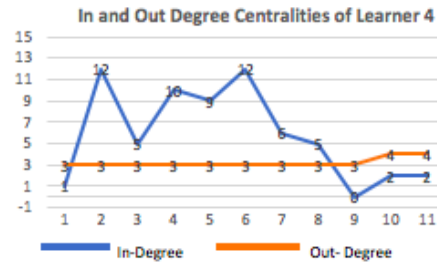


Figure 1. The in- and out-degree centralities of learner #4 from week 1 to week 11.

Correlation analysis found temporal dispersion to be positively correlated with the closeness centrality, indicating that, on average the more spread out learners' annotations are throughout the week, the easier they could reach their peers in the reply network. The timing of annotation was negatively correlated with in-degree and closeness centralities, implying early starters (coded as 0) occupied more central positions (see Table 1).

Table 1. Correlation coefficients among learner temporal measures, local centralities, and densities.

Temporal Variables	In-degree Centrality	Out-degree Centrality	Closeness Centrality	Ego Density
Temporal Dispersion	0.39 ($p=.10$)	0.24 ($p=.07$)	0.26($p=.04$)	0.11 ($p=.08$)
Timing of Annotation	-0.48 ($p=.07$)	0.17 ($p=.50$)	-0.53 ($p=.06$)	0.20 ($p=.40$)

4 DISCUSSION AND CONCLUSIONS

We attempted to combine temporal and social network analysis, slicing the learner interaction network by week and computing two temporal measures to characterize temporality of learners' participation. Positive correlations between temporal dispersion and centralities indicated that the more time a learner spends throughout the week checking on annotations and responses, the more prominent their network position is in social annotation activities. While this study was focused on one small online class, it highlights the importance of integrating social and temporal dimensions of learning. Future research is clearly needed to investigate the temporal unfolding of social interaction in similar learning contexts.

REFERENCES

- Chen, B., Knight, S., & Wise, A. F. (2018). Critical issues in designing and implementing temporal analytics. *Journal of Learning Analytics*, 5(1), 1–9. <https://doi.org/10.18608/jla.2018.53.1>
- Mendenhall, A. T. E. (2010). Fostering the development of critical thinking skills, and reading comprehension of undergraduates using a Web 2.0 tool coupled with a learning system. *Interactive Learning Environments*, 18(3), 263-276. doi: 10.1080/10494820.2010.500537
- Zurita, G., & Nussbaum, M. (2004b). "Computer-Supported Collaborative Learning using Wirelessly Interconnected Handheld Computers." *Computers & Education*. 42: 289–314.

Barriers and Hurdles to the Publication of Learning Analytics Data

Katarzyna Biernacka

Humboldt-Universität zu Berlin
biernack@hu-berlin.de

Niels Pinkwart

Humboldt-Universität zu Berlin
pinkwart@hu-berlin.de

ABSTRACT: The publication of research data involves many legal areas and their respective legal complexity (Hartmann, 2019). The most common denominator, however, is the data protection law. It requires ‘de-identification’ to protect privacy of the research participants. In the case of learning analytics data, anonymization can lead to the loss of some relevant information as the identifying data can improve the student’s learning experience, however, it is important to publish research data so as to keep the research transparent and replicable in terms of Open Science. This paper provides preliminary results from the guided interview study on the presumed conflict between publishing learning analytics data and the issues of privacy protection.

Keywords: research data, open data, research integrity, data protection

1 RESEARCH PROBLEM AND PURPOSE

Ideally, research data should be open and made publicly available (ALLEA, 2017) in terms of Research Integrity and Open Science. Open Data is one of the four pillars of Open Science and can help to reduce misconduct, facilitate replication (i.e. reproducibility), and support further research (e.g. meta-analyses). Several scientific articles already show the benefits of it, e.g. increased citation frequency (H. A. Piwowar, Day, & Fridsma, 2007; Heather A. Piwowar & Vision, 2013). In addition, more and more funders, publishers and research institutions are demanding the exchange of data (European Commission, 2016; Jones, Grant, & Hrynaszkiewicz, 2019). And yet, learning analytics data seems not to be sufficiently published. Since learning analytics systems collect amongst other things personal data about individuals, they are also subject to data protection laws and regulations. However, the unfamiliarity with the EU General Data Protection Regulation (GDPR) and other national legislation or related EU measures create a high level of uncertainty among the scientific community: Who is the copyright holder? Is the anonymized data still useful? What if someone ‘steals’ my data? Is the improvement of the overall learning experience a valid reason to record and share personal data to facilitate research on the topic? Is the data allowed to be published at all? The extent to which principles and standards apply to research data with varying degrees of person-related information remains largely unclear. It requires new reflection on the compatibility of good research and good data protection.

In the research presented in this paper the presumed conflict between publishing of learning analytics data and the issues of data protection in Germany is investigated. Our research questions are:

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

- What are the biggest concerns when publishing learning analytics data?
- How does the publication of learning analytics data compare to other disciplines?
- Does GDPR prevent the publication of learning analytics data in Germany?

2 METHODOLOGY

In the first phase of the research the various barriers to the publication of learning analytics data were investigated. For this purpose, guided interviews were conducted with scientists in order to assess their handling of the research data. The guided interview gave them the possibility to talk about their personal experience with repositories, legal issues and other obstacles while publishing their data.

For the qualitative evaluation of the personal guided interviews the Grounded Theory methodology was chosen based on the guidelines of Charmaz (Charmaz, 2006). The transcribed interviews have been coded with the categorization of segments of the data with the inductive coding approach.

As practices around research data vary between disciplines, we make a comparison between three different disciplines in order to exclude or confirm discipline-specific publishing phenomenon. In addition to learning analytics, medicine and climate impact research were chosen. The considered disciplines show large variations in the type of research data in terms of data sensitivity.

3 PRELIMINARY RESULTS

At the point of submitting the research paper 12 guided interviews with scientist from Germany had been conducted: 4 on learning analytics, 4 on climate impact research and 4 on medicine. Both, junior (2-4 years of experience) and senior scientist (more than 5 years of experience), from 10 different institutions were interviewed.

In the first interviews conducted with scientists in learning analytics, the following codes for the barriers to the publication of learning analytics data have emerged¹: *balancing privacy and openness*, *unclear responsibilities*, *personal and/or sensitive data*, non-visible value, complexity of the publication process, *anonymization (no complete security)*, *anonymization (loss of information)*, *anonymization (conducting the anonymization process)*, *uncertainty what is allowed*, costs and competition. These can be seen in the following example:

[...] because – as far as we know – it is not possible simply to make the data available, because there is also a great amount of personal data available. [...] [Anonymization] is a problem. [...] we need it [the personal information] to take further steps, e.g. to perform personalization. (*Junior Scientist, Learning Analytics*)

In the guided interviews with scientist in climate impact research two new codes emerged: time effort and not established in community. This was in addition to the already mentioned codes

¹ Italicized codes are direct or indirect impact of the implementation of the GDPR in Germany.

uncertainty what is allowed, non-visible value, complexity of the publication process and competition. Compare with the following example:

Until recently, it was not really a topic there [in the community] and there was no discussion about it. [...] and now that people are talking about it, some people are worried to give away their treasure. (*Senior Scientist, Climate Impact Research*)

In medicine five barriers were mentioned: *personal and/or sensitive data*, complexity of the publication process, competition, non-visible value and time effort. From the context of the guided interviews one could also see that it is not a common approach in the community (not established in community):

This is just now coming up [the publication of research data] and then perhaps in the next step, there is also the problem of data interpretation, perhaps, which can then be done by those who can use the data. (*Senior Scientist, Medicine*)

Seven of the 13 mentioned barriers can be seen as direct or indirect impact of the implementation of the GDPR in Germany and all of those were mentioned by scientist in learning analytics. The *uncertainty what is allowed* was mentioned by scientist in climate impact research too, although the data produced by this discipline is mostly neither personal nor sensitive. In contrast in medicine, personal and/or sensitive data is a problem by definition. The emerged codes indicate therefore, that the EU General Data Protection Regulation from May 2018 may be a factor that intensifies the conflict between sharing research data on the one hand and data privacy on the other.

4 OUTLOOK

As different legal regulations and cultural factors lead to different starting situations, three more countries will be examined in the next step: China, India and Peru. Furthermore, the theories obtained from the guided interviews will be verified by means of widely disseminated online surveys. From this, guidelines, recommendations and best practices as well as generalizable strategies for the publication of learning analytics research data will be derived and tested.

REFERENCES

- ALLEA. (2017) The European Code of Conduct for Research Integrity. In, (pp. 15). Berlin: ALLEA - All European Academies.
- Charmaz, K. (2006). *Constructing grounded theory: a practical guide through qualitative analysis* (Reprinted ed.). London u.a.: Sage.
- European Commission. (2016). H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020. In (pp. 12): European Commission.
- Hartmann, T. (2019). Rechtsfragen. Institutioneller Rahmen und Handlungsoptionen für universitäres FDM. In. Frankfurt (Oder): Europa-Universität Viadrina Frankfurt (Oder).
- Jones, L., Grant, R., & Hrynaszkiewicz, I. (2019). Implementing publisher policies that inform, support and encourage authors to share data: two case studies. *Insights the UKSG journal*, 32, 11. doi:10.1629/uksg.463
- Piowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS One*, 2(3), 5. doi:10.1371/journal.pone.0000308
- Piowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*(1:e175), 25. doi:10.7717/peerj.175

Physiology-Aware Learning Analytics Using Pedagogical Agents

Melanie Bleck, Nguyen-Thinh Le, Niels Pinkwart

Humboldt-Universität zu Berlin

{nguyen-thinh.le, niels.pinkwart}@hu-berlin.de, mail@melanie-bleck.de

ABSTRACT: Learning analytics applications consider not only the cognitive dimension, but also the physiological dimension of the learner. This paper describes a learning analytics approach that focuses on alerting the critical stress level of the learner using a pedagogical agent. For that purpose, an existing pedagogical agent was expanded by a software component, which analyses heart rate variability data to determine the cognitive load of a user and to offer support with stress reduction. The evaluation study with the physiology-aware pedagogical agent showed an improvement of learning and a reduction of stress.

Keywords: Physiological computing, Heart rate variability, pedagogical agent

1 INTRODUCTION

One of the factors that may affect learning performance is stress (Li et al., 2017). Thus, detecting and measuring stress that occurs while learning could be used to enhance learning analytics applications. In addition to proposals to different observation techniques, e.g., facial detection and video monitoring (D’Mello, 2017), the monitoring of physiological parameters like Heart Rate Variability (HRV) is considered a relevant indicator for the stress detection (Zangroniz et al., 2018). However, handling physiological data, to what extent they can be used to analyze excessive cognitive demands and how it can be utilized in a learning analytics context are still a research gap. The research question to be investigated in this paper is how HRV data can be used by pedagogical agents to determine the stress level of the learner and to alert the learner in a learning situation.

2 METHODOLOGY

In order to investigate the specified research question, the functionality of the web-based pedagogical agent LIZA (Le & Wartschinski, 2018) aimed at improving the decision making and reasoning of the user, was extended through three different parts. The first component provides a solution to generate, save and process the HRV data, the second one analyses the data regarding stress and the third one adapts the learning situation through selected stress reduction strategies. To determine the effectiveness and benefits of the approach, the adjusted pedagogical agent was evaluated.

To use HRV parameters to determine stress level, the generation of data has to be ensured. A technical solution was provided by the wristband E4 of Empatica. The integrated photoplethmography sensor in Empatica E4 was utilized to determine the heart rate and the time interval between two consecutive heartbeats (NN Interval) (Empatica Inc., 2016). If a specific time interval is requested by the pedagogical agent, the suitable NN Intervals will be selected on the basis of the time stamp and the root mean square of successive differences in the heart rate (RMSSD) of these values will be determined. RMSSD was chosen as a metric for HRV because of the

recommendation as an indicator for cognitive load in short term measurements (less than 5 minutes) by AWMF (Sammito et al., 2014).

First, the user is requested to apply and activate the wristband und start the mobile transmission application in a preparation phase. Since there are no generally accepted threshold values to determine the degree of a mental load of a person, a series of individual measurements has to be done (Sammito et al., 2014). But that alone does not provide enough information to automatically identify an overload during a certain task. A range of cognitive load has to be identified and a specific threshold for adapting a learning situation has to be defined. Because of that, a phase was added in the pedagogical intervention process, in which two different levels of stress are induced, the RMSSDs are calculated accordingly and then used as an indicator for different stress levels later on. Arithmetic tasks were chosen after an analysis of induction methods for cognitive load of several research papers and they have been used widely to generate moderate stress level (Schneider et al., 2003). In two arithmetic stress tests, with different levels of difficulty, the user had to subtract a random value from a certain number consecutively for five minutes. The level of difficulty was altered via the time limit for solving the equation, the number of digits of the random value and the value of the start minuend. Furthermore, a competition situation was created by requesting to beat LIZA in the number of correct answers under certain conditions in the second test. With that, the first test provided the RMSSD value for a moderate load, the second, which is designed with a higher difficulty level, for an overload. Certainly, a range of cognitive load could be defined in that way, but a specific threshold could not be derived from this data. Considering that an excessive load can result in a decrease of motivation and an abort of the learning in the long term, the learning situation has to be adapted before such a scenario materializes. Another factor is that an adaption of a learning situation based on adaptive stress reduction strategies will interrupt the process itself. So, it should be carried out as seldom as possible but also as often as necessary. After a preliminary empirical test with a threshold at 50% of the individually defined stress range, which triggered an intervention nearly every time LIZA was used, the threshold was increased to 2/3 and the intervention could be reduced to 40% of the cases. If the current RMSSD falls below the threshold after a specific time, LIZA offers assistance in reducing the stress level through stress reduction strategies. Among different strategies (e.g., mindfulness-based stress reduction, autogenic training), two methods were selected that seemed appropriate for the learning environment of a pedagogical agent. The first one distracts the user by telling jokes, the second one shows a video with relaxing content. The user decides whether it is necessary to start the coping process and how long the strategies are used until the stress level is significantly reduced.

3 EVALUATION

The goal of the evaluation study we conducted was to determine the effectiveness and benefits of the pedagogical agent that was extended with the capability of measuring HRV and detecting critical stress level of learners. Amongst others, the following hypotheses were examined: 1) The stress reduction strategies lead to a relaxation of learners; 2) The RMSSD is a suitable indicator for cognitive load; 3) The adaption of the learning process affects the learning performance. To examine these hypotheses, a pre- and posttest were performed. For the study, 34 participants (10 males, 24 females) aged between 21 and 59 (mean 31 ± 11 years) were acquired and assigned to test- or control condition randomly. The test was conducted in a quiet environment under supervision. Every

participant was asked to use the pedagogical agent independently. In the learning phase 4 tasks had to be solved by the participants, where every task covered a different problem of reasoning. After that, the RMSSD was calculated for the time frame of the first task block and compared with the previously determined stress limits. For the test group, a stress reduction phase followed if the current RMSSD fell below the threshold. Both groups continued with the posttest that required all participants to solve 4 tasks again. The reasoning problems of the pretest and posttest were the same, but the tasks were different. In the end, the participants got an evaluation of how successful the tasks were solved. For every part of the process, the RMSSD was calculated so that the development of the indicator could be retraced. In addition, the participants had to self-assess their current state of cognitive load with a short questionnaire (KAB) (Wagner, 2012).

The first hypothesis, which covers whether stress reduction leads to the relaxation of the learner, could be partly confirmed. In the self-assessments, 90% of the participants confirmed the effect of the applied stress reduction strategies. But only in nearly 50% of the cases, the RMSSD also fell below the threshold. Possible reasons for that could be deficits in stress limit determination, insufficient choice of strategies or application time. Concerning the adequacy of the RMSSD as an indicator for stress, we found weak negative correlation ($r = -0,04$) between the RMSSD values and the KAB-Index which did not reach the level of statistical significance ($p = 0.67$). One reason could be the error-prone self-assessment like Picard points out (Picard, 2003). Finally, the test group showed a significantly higher improvement in terms of learning gains ($p = 0.07$ at a significance level of 10%) as compared to the control group.

REFERENCES

- Empatica Inc. (Hrsg.). (2016). *Utilizing the PPG/BVP Signal*. Accessed 01/11/2019 <https://support.empatica.com/hc/en-us/articles/204954639-Utilizing-the-PPG-BVP-signal>
- Li, Q., Xue, Y., et al. (2017). Analyzing and Identifying Teens Stressful Periods and Stressor Events from a Microblog. *IEEE Journal of Biomedical and Health Informatics*, 21(5), 1434-1448.
- Picard, R. W. (2003). Affective Computing: Challenges. *Int. J. of Human-Computer Studies*, 59, 55-64.
- Sammito, S., Thielmann, B., et al. (2014). S2k-Leitlinie: Nutzung der Herzfrequenz und der Herzfrequenzvariabilität in der Arbeitsmedizin und Arbeitswissenschaft. *AWMF online*.
- Schneider, M., Jacobs, D. W., et al. (2003). Cardiovascular Hemodynamic Response to Repeated Mental Stress in Normotensive Subjects at Genetic Risk of Hypertension: Evidence of Enhanced Reactivity, Blunted Adaptation and Delayed Recovery. *J. of Human Hypertension*, 17(12), 829-840.
- Sidney D'Mello, A. K. (2017). The Affective Computing Approach to Affect Measurement. *Emotion Review*. doi:10.1177/1754073917696583
- Wagner, M. (2012). Der Einfluss emotionaler, kognitiver und körperlicher Beanspruchung auf die Herzrate bei Jugendlichen: Eine Laborstudie. *Ph.D. Dissertation, Universität Ulm, Germany*.
- Le, N. T. & Wartschinski, L. (2018). A Cognitive Assistant for Improving Human Reasoning Skills. *International Journal of Human-Computer Studies*. 10.1016/j.ijhcs.2018.02.005.
- Zangroniz, R., Martinez-Rodrigo, A. et al. (2018). Estimation of Mental Distress from Photoplethysmography. *Applied Science*, 8(69), 1-15.

Exploring Writing Achievement and Genre in Postsecondary Writing

Jill Burstein

Educational Testing Service
jburstein@ets.org

Daniel McCaffrey

Educational Testing Service
dmccaffrey@ets.org

Norbert Elliot

University of South Florida
nelliot3@usf.edu

Beata Beigman Klebanov

Educational Testing Service
bbeigmanklebanov@ets.org

ABSTRACT: Writing achievement is a complex skill set as characterized by the sociocognitive writing framework, including writing domain knowledge (e.g., sentence structure), general cognitive skills (e.g., critical thinking) and intra- (e.g., interest) and interpersonal (e.g., collaboration) subfactors. During students' postsecondary careers, they need to write in different genres. Yet, we have limited understanding about the contribution of genre mastery to students' writing achievement which can affect their broader success (e.g., GPA). Partnering with six, diverse 4-year universities, we collected student responses to a *standardized* writing assessment and *authentic* course writing assignments which were coded for *genre* as: *standardized*, *persuasive*, *inform/explore*, and *reflective*. Using automated writing evaluation, we extracted approximately 50 linguistic features (e.g., vocabulary usage) from the 1,426 writing samples. We present findings for genre-based feature distributions, cross-genre correlations, and implications for postsecondary writing education.

Keywords: natural language processing, writing analytics, higher education

1 INTRODUCTION

Writing achievement is a complex skill set as characterized by the sociocognitive writing model (Flower, 1994; Hayes, 2012). The model considers multiple subfactors, including writing domain knowledge (e.g., sentence structure), general cognitive skills (e.g., critical thinking), and intra- (e.g., interest) and interpersonal (e.g., collaboration) subfactors. Postsecondary writing achievement studies are needed to critically examine how students apply and develop their writing domain knowledge in different genres, since writing achievement may affect broader success measure, e.g., GPA (Burstein, McCaffrey, Beigman Klebanov, Ling, & Holtzman, 2019). Such studies have typically examined expository essay writing genre (Allen, Snow, Crossley, Jackson, & McNamara, 2014; MacArthur, Traga Philippakos, May, & Compello, 2019). Burstein, et al. (2019) used standardized writing assessment and coursework writing to examine relationships between automated writing evaluation (AWE) features and academic success measures (e.g., GPA); yet, genre was not studied. Our study compares writing subconstruct features in student writing as captured by state-of-the-art AWE technology (*withheld for anonymity*) between genres.

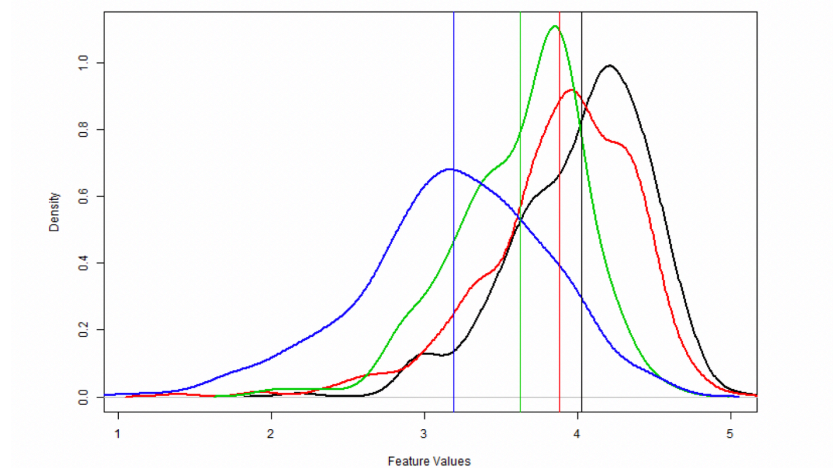


Figure 1: *Standardized* writing (blue) has lower sentence variety values than *reflective* (green), *persuasive* (red), or *informative/exploratory* (black) sentence variety values.

2 METHODS

2.1 Data

At six diverse, 4-year partner universities, 735 students participated. We collected 1,426 writing samples. A subset of students completed a *timed, standardized writing assessment* requiring an argumentative essay ($n=366$). A partially overlapping subset of students ($n=435$) submitted coursework writing ($n=1060$) from one course in which their instructor had agreed to participate for the study. Courses were primarily first-year English courses, but also included Biology, Business, Exercise Science, History, and Sociology courses. Data are available here: <https://github.com/EducationalTestingService/ies-writing-achievement-study-data>.

2.2 Genre Annotation

Three research assistants annotated the writing samples with four genre labels. All *timed, standardized writing assessment responses* were labeled as “*standardized*” (**S**) and *coursework writing* was coded as one of “*persuasive*” (**P**) (33%), “*informative/exploratory*” (**IE**) (47%), “*reflective*” (**R**) (14%), or “*other*” (5%), using an annotation protocol developed for the study. “*Other*” assignments did not align with the 3 *coursework* genres, and are not included in this discussion.

2.3 Data Analysis, Results & Discussion

Using AWE, we extracted about 50 linguistic features from the standardized and coursework writing samples. The feature set represented six writing subconstructs: *vocabulary usage*, *argumentation*, *organization & development*, *English conventions*, *sentence structure*, and *personal reflection*.

Feature Density & Genre. Using visual comparisons of smoothed density plots, and the Kolmogorov-Smirnov test for differences in the distributions for the different subconstructs, we observed statistically-significant ($p < 0.001$), genre-based differences in AWE feature distributions. For instance, more pronouns (i.e., *personal reflection*) were observed in *reflective* writing than *standardized*, *persuasive*, or *informative/exploratory* writing. Analyses suggested that *standardized* writing (a)

contained less sentence variety (i.e., *sentence structure*) than the coursework genres (e.g., Figure 1), (b) used less sophisticated vocabulary (i.e., *vocabulary usage*) than coursework genres, and (c) tended to discuss one longer topic, (i.e., *development*), whereas coursework genres contained more topic variety.

Cross-Genre Correlations. We generated ‘*subconstruct scores*’ for the six writing subconstructs. *Subconstruct scores* were equal to the average of the AWE feature values for the features in each subconstruct. Feature values were standardized to have a mean zero and standard deviation 1 prior to averaging. Six factor scores were assigned to all writing samples. We ran cross-genre correlations to examine relationships between the *subconstruct scores* for writing samples in each genre pair (e.g., R/IE). Coursework genre pairs had the highest correlations for *vocabulary usage* (0.35 for P/IE, and 0.31 for IE/R), *English conventions* (0.33 for P/IE and 0.35 for IE/R), and *sentence structure* (0.30 for IE/R) subconstructs. Correlations between S and the *coursework genres* all fell below 0.30.

Implications. Findings from both analyses suggested differences in students’ application of writing features across genres. Offering opportunities for students to practice writing in different genres can provide instructors and institutions with a more comprehensive picture of students’ writing domain knowledge (i.e., writing feature use) and writing achievement. The findings illustrate the limitations of observable writing domain knowledge from *single-genre* standardized writing assessments.

ACKNOWLEDGMENTS

Research presented in this paper was funded by the Institute of Education Science, United States Department of Education, Award Number R305A160115 any opinions, findings, conclusions, or recommendations are those of the authors and do not necessarily reflect the views of the IES. This project would not have been possible without the support of Hillary Molloy, Patrick Houghton and Zydrune Mladineo who led and performed the genre annotation.

REFERENCES

- Allen, L. K., Snow, E. L., & Crossley, S. A., Jackson, G. T., & McNamara, D. S. (2014). Reading components and their relation to the writing process. *L’année psychologique/Topics in Cognitive Psychology*, 114, 663-691.
- Burstein, J., McCaffrey, D., Beigman Klebanov, B., Ling, G. & Holtzman, S. (2019). Exploring Writing Analytics and Postsecondary Success Indicators. In *Companion Proceedings 9th International Conference on Learning Analytics & Knowledge* (pp. 213-214). Retrieved from http://solaresearch.org/uploads/LAK19_Companion_Proceedings.pdf
- Flower, L. (1994). *The construction of negotiated meaning: A social cognitive theory of writing*. Carbondale and Edwardsville, IL: Southern Illinois University Press.
- Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication*, 29, 369–88. <https://doi.org/10.1177%2F0741088312451260>
- MacArthur, C. A., Jennings, A., Philippakos, Z. A. (2019). Which linguistic features predict quality of argumentative writing for college basic writers, and how do those features change with instruction? *Reading and Writing: An Interdisciplinary Journal*, 32(6), 1553-1574. <https://doi.org/10.1007/s11145-018-9853-6>.

Assessing Risk in Learning Analytics Projects

Henrique Chevreux¹, Valeria Henríquez¹, Eliana Scheihing¹, Pedro Muñoz-Merino², Tinne De Laet³, Mar Pérez-Sanagustín⁴, Isabel Hilliger⁴, Jorge Maldonado-Mahauad⁵, Paola Pesantez⁵, and Margarita Ortiz⁶

Universidad Austral de Chile¹, Universidad Carlos III de Madrid², KU Leuven³, PUC Chile⁴, Universidad de Cuenca⁵, Escuela Superior Politécnica del Litoral⁶
{henrique.chevreux, valeria.henriquez, escheihi}@inf.uach.cl

ABSTRACT: Learning Analytics (LA) enables leaders to improve teaching, learning, organizational efficiency, and decision making. Nonetheless, LA initiatives often have difficulty to move out of prototype into real educational practice. As an emerging multidisciplinary field, we wonder how much its challenges are similar to other more mature related fields. This work assesses to which extent its risks are related to another well-established field as Enterprise Resource Planning (ERP) projects. Our findings show that risk factors and categories in ERP apply to LA projects and their top risks are considered very similar by LA experts. Therefore this work can help the LA community in the search for strategies to sustainable adoption.

Keywords: Learning Analytics, risk management, process reengineering, adoption strategies

1 INTRODUCTION AND RELATED WORK

It has proven to be challenging to create scalable implementations of Learning Analytics (LA) in authentic contexts that go beyond a particular course or setting (De Laet et al., 2018; Ferguson et al., 2014). By nature, LA projects usually involve the development and/or adoption of software. As software engineering suffers from the habit of paying too little attention to how other engineers do engineering (Denning & Riehle, 2009), LA projects may suffer from the habit of paying little attention to how projects are implemented in practice in more mature fields. For decades the Enterprise Resource Planning (ERP) system market has been one of the fastest-growing markets in the software industry (Huang et al., 2004). Although LA and ERP projects pursue different goals, both share similar characteristics, such as: operational and managerial processes change simultaneously; needed coordination among different departments and different hierarchical levels; provide opportunities for or require organizational process reengineering. These shared complexities suggest that techniques from the more mature ERP field could be applied, at least partially, to increase the probabilities of the success of LA projects. This work validates their similarities in one of the most important engineering tasks, risk management (Denning & Riehle, 2009). Specifically, we investigate the following research question (RQ): **How applicable are ERP risk factors to LA projects?**

2 MATERIALS AND METHODS

We developed an anonymous survey based on the risk factors identified and prioritized with Delphi and AHP methods in “*Assessing risk in ERP projects: identify and prioritize the factors*” (Huang et al., 2004). We considered its 6 categories and 28 factors as a good base to develop the questionnaire to answer our RQ. The survey is organized through 3 sections: (1) demographic questions to characterize the participants; (2) questions to assess how important are each of the six categories for a successful LA project; (3) questions per category to assess how much each factor risks LA projects. For the last two sections, we used a Likert scale with 7 items and we also provided open-ended questions to add

new factors. The final survey is composed of 39 mandatory items (5 demographic + 6 categories + 28 factors) plus 7 optional per person. We sent the survey in October of 2019 by email to all 46 LALA (Building Capacity to Use Learning Analytics to Improve Higher Education in Latin America - <https://www.lalaproject.org>) members. We received 15 responses, a 33% response rate. To avoid bias, we didn't mention the source of the risk factors and adapted keywords that could give it away indirectly, like *ERP/enterprise/business* (changed to *LA/organization/educational*).

3 RESULTS

Table 1 shows that all participants have experience in LA, most are researchers and/or professors and have more than two years of experience in LA projects.

Table 1: Sample demographic.

Characteristics	Total	Characteristics	Total
Current position (could select more than one)		Work experience	
Manager	5	1-5 years	2
Professor	9	6-10 years	6
Researcher	12	11-15 years	5
Engineer	2	16 or above	2
Other	0	Years of work experience in LA	
Organization size		0 < years of experience ≤ 2	3
Under 1000 employees	5	2 < years of experience ≤ 4	9
1000 or above	10	More than 4 years	3

Regarding the **RQ (How applicable are ERP risk factors to LA projects?)**, the results obtained in the survey show that both the categories and risk factors in ERP apply to LA projects. All 6 categories and 22 out of 28 factors averaged more than 5 points out of 7. No factor averaged less than 4.5 points. In both cases, the top categories were “*User involvement and training*” and “*Project management and control*”. Table 2 shows that six of the top ten LA risk factors are in the top ten risks of ERP projects. Four of the first five factors appear on both (three in the same position). Among the top five, the biggest differences are the entry of the “*Lack of integration between organization-wide systems*” factor in the LA top list and the descending of the “*Lack of senior manager commitment to project*” factor to the 9th position. Both may be explained because the educational environment is usually less hierarchical than the business one. Hence in LA convincing and cooperation probably yields more results than “*senior management commitment*”. This also would explain why all the top three risk factors of LA projects are explicitly related to the ‘user’ of the system.

Table 2: Top ten risk factors in LA projects versus top risk factors in ERP.

Top ten risk factors in LA projects	Priority in LA	Priority in ERP
Fail to get user support	1	4
Ineffective communication with user	2	2
Insufficient training of end-user	3	3
Lack of integration between organization-wide systems	4	-
Lack of effective project management methodology	5	5
Extent of change	6	-
Developing wrong functions and wrong user interface	7	-
Lack of appropriate experience of the user representatives	8	-
Lack of senior manager commitment to project	9	1
Attempting to link legacy systems	10	6

The figure at <https://ibb.co/jzbjKS6> shows the result details by categories, ordered by importance in LA projects. The numbers in the circles show the corresponding importance in ERP. Four participants

added a total of ten risk factors to take into consideration. It gives us the opportunity of exploring in the future to what extent they are relevant to the LA community. These added factors are illustrated by boxes with dotted lines.

4 CONCLUSIONS

Our work informs that risk factors on ERP apply to LA projects. Therefore we call the LA community to put efforts to mitigate those risks, increasing the chances of sustainable adoption. In particular, one participant mentions: *"I notice that I identify most of the above as high risk, while many of them are currently ignored in most LA guidelines and frameworks."* Another call may be for the integration of more professionals with practical engineering experience. Academics and learning analytics tools are usually developed inside research departments (Chevreux et al., 2019) and many software developers have been raised in a research tradition, not an engineering tradition (Denning & Riehle, 2009). The user-centered methodologies are learned more in the industry or are self-taught, i.e., not pervasive yet to the research environment (Vargas et al., 2018). This work is incipient, and findings can not be generalized, but it suggests that if risk factors of both types of projects are related, then solutions may be related too. Future work can go deeper on the relationship between LA and ERP fields and evaluate the similarities of risk factors and well-tested solutions to address them. We hope it serves to inspire the search for solutions in other related fields as well. As an emerging discipline, LA initiatives may benefit from standing on the shoulders of more mature similar fields, like ERP and process reengineering projects.

5 ACKNOWLEDGEMENTS

Work funded by the LALA project (grant no. 586120-EPP-1-2017-1-ES-EPPKA2-CBHE-JP). The LALA project has been funded with support from the European Commission. This publication reflects the views only of the authors, and the Commission and the Agency cannot be held responsible for any use which may be made of the information contained therein.

REFERENCES

- Huang, S. M., Chang, I. C., Li, S. H., & Lin, M. T. (2004). Assessing risk in ERP projects: identify and prioritize the factors. *Industrial management & data systems*, 104(8), 681-688.
- De Laet, T., et al. (2018). Involving Stakeholders in Learning Analytics: Opportunity or Threat for Learning Analytics at Scale?. In *Proceedings of the 8th International Learning Analytics & Knowledge Conference (LAK'18)*. Society for Learning Analytics Research, 602-606.
- Chevreux, H., Henríquez, V., Guerra, J., & Scheihing, E. (2019, September). Agile Development of Learning Analytics Tools in a Rigid Environment like a University: Benefits, Challenges and Strategies. In *European Conference on Technology Enhanced Learning* (pp. 705-708).
- Ferguson, R., Clow, D., Macfadyen, L., Essa, A., Dawson, S., & Alexander, S. (2014).: Setting learning analytics in context. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge - LAK 14* (pp. 251-253)
- Vargas, T., Chevreux, H., Henriquez, V., & Lima, H. (2018). Continuous Integration in Chile: Knowledge, Perception, and Interest. *Jornadas Iberoamericanas de Ingeniería de Software e Ingeniería del Conocimiento - JIISIC 2018* (pp. 247-259)
- Denning, P. J., & Riehle, R. D. (2009). The profession of IT. Is software engineering engineering?. *Communications of the ACM*, 52(3), 24-26.

EduBERT: Pretrained Deep Language Models for Learning Analytics

Benjamin Clavié, Kobi Gal
The University of Edinburgh
{Benjamin.clavie;kgal}@ed.ac.uk

ABSTRACT: The use of large pretrained neural networks to create contextualized word embeddings has drastically improved performance on several natural language processing (NLP) tasks. These computationally expensive models have begun to be applied to domain-specific NLP tasks such as re-hospitalization prediction from clinical notes. This paper demonstrates that using large pretrained models produces excellent results on common learning analytics tasks. Pre-training deep language models using student forum data from a wide array of online courses improves performance beyond the state of the art on three text classification tasks. We also show that a smaller, distilled version of our model produces the best results on two of the three tasks while limiting computational cost. We make both models available to the research community at large.¹

Keywords: MOOC Forums, Natural Language Processing, Online Learning, Text Classification, Pretrained Models, Deep Learning

INTRODUCTION

In the past year, the field of Natural Language Processing (NLP) has seen the rise of pretrained language models such as as ELMo (Peters et al., 2018), ULMFiT (Howard and Ruder, 2018) and BERT (Devlin et al., 2019). These approaches train a deep-learning language model on large volumes of unlabelled text, before fine-tuning it for particular NLP tasks. Such models achieved state-of-the-art performance on tasks ranging from sentiment classification to question answering (Devlin et al., 2019).

The benefit of these models has also been demonstrated in specialized NLP domains. BioBERT (Lee et al., 2019), a version of BERT trained exclusively on biomedical text, was able to significantly increase performance on biomedical named entity recognition. Further refining this model on clinical text produced an increase in performance in medical natural language inference (Alsentzer et al. 2019).

While large pretrained models offer significantly increased performance, they come with their own constraints. The number of parameters in the classic BERT-base model exceeds 100 million. As such, the computational cost can be prohibitively high at both training and prediction time (Devlin et al., 2019). Recent work has addressed this challenge by ‘distilling’ the models, training smaller versions of BERT which reduce the number of parameters by 40% while retaining more than 95% of the full model performance and even outperforming it on two out of eleven GLUE tasks (Sanh et al., 2019).

This paper shows that using pretrained models in learning analytics holds great potential for advancing the field. We apply the BERT approach to the following three previously explored LAK tasks using MOOC forum data (Wei et al., 2017): Confusion detection, urgency of teacher intervention and

¹ Available at <https://github.com/bclavie/edubert>

sentimentality classification. In all three tasks, we are able to improve performance past the state of the art.

METHOD

Data: We trained the language model on a large unannotated data set from two sources: student forum data from the Stanford MOOCPosts dataset (Agrawal and Paepcke, 2014) which includes about 30,000 forum posts from 11 courses; and forum data from multiple instances of 18 courses from large public universities in the UK and USA. In total, this dataset is comprised of more than 12 million tokens.

The data used for the classification tasks was from the same Stanford MOOCPosts dataset. The posts are annotated by domain experts and given scores for **urgency** for the post to receive a response from an instructor as well as **sentimentality** and **confusion** expressed by the student and. Scores are given on a Likert scale from 1 (low) to 7 (high). The annotation process is detailed in Agrawal and Paepcke (2014). It aimed to produce a gold set by having three human experts annotating each example and selecting the annotator combinations producing the highest agreement.

Language Models: We constructed two models, EduBERT and EduDistilBERT, which respectively refine BERT-base and DistilBERT (Sanh et al., 2019), both of which were trained on general domain text from books and Wikipedia (Devlin et al., 2019). Both models are initialized from their base model and fine-tuned on educational data, using the Transformers library (Wolf et al., 2019). The fine-tuning step allows the model to better capture how words are used in an educational context.

Training of the models was performed on a Titan X GPU. We set the maximum input sequence length to the default value (512); the learning rate to $5e-5$; the batch size (# of input sequences processed at one time) to 8 for EduBERT and 16 for EduDistilBERT. Best performance was achieved after 5 epochs.

Classification Tasks: To encourage easily comparable results, we evaluated the models on three well-explored classification tasks on the StanfordMOOC dataset. Following previous work by Guo et al. (2019), we split the data into a 2/3 training set and 1/3 test set and consider a post to express sentiment, urgency or confusion if and only if its respective score is ≥ 4 .

We compare between the four classifiers BERT-base, DistilBERT, EduBERT and EduDistilBERT. We evaluated multiple sets of parameters. Best results for these tasks were achieved with the following parameters: two learning epochs, maximal sequence length of 300 (BERT-base, EDUBERT) and 512 for the distilled models, all other parameter values were equal to the ones used for pre-training.

RESULTS & DISCUSSIONS

Table 1 compares EduBERT, EduDistilBERT to their base versions, as well as the state-of-the-art (SoA) for urgency detection (Guo et al. 2019). The table shows that all pretraining approaches outperformed the SoA for F1 and weighted F1 measures, with our distilled model EduDistilBERT achieving the best overall performance. **Table 2** compares all of the models for all three tasks to the SoA using the same measures of accuracy as Wei et al. (2017). Again, all the pretraining approaches outperform the SoA. EduDistilBERT obtains the best results on both urgency and confusion prediction while EduBERT performs the best for sentimentality classification. However, EduDistilBERT has a lower memory footprint and is noticeably faster at inference time, allowing for a 30% speedup.

The ability to detect the urgency and confusion expressed by learners in MOOCs forum is an essential step towards enhancing the learning experience provided by MOOCs. Indeed, online courses

have extremely high student drop-out rate (Onah et al., 2014). We believe that tools able to efficiently detect posts expressing confusion or urgency could help instructors navigate the large amount of forum posts they may encounter and address students' concerns before they become discouraged and abandon the course.

Table 1: Performance metrics for urgency prediction

	Non-urgent			Urgent			
	Recall	Precision	F1	Recall	Precision	F1	Weighted F1
EduDistilBERT (Ours)	0.949	0.954	0.952	0.835	0.819	0.827	0.925
DistilBERT	0.946	0.953	0.950	0.833	0.810	0.821	0.921
EduBERT (Ours)	0.950	0.950	0.950	0.822	0.820	0.821	0.922
BERT-base	0.956	0.944	0.950	0.794	0.835	0.814	0.920
Guo et al. (2019)	0.954	0.948	0.951	0.772	0.834	0.801	0.918

Table 2: Accuracy measures on the three tasks

	Confusion	Sentiment	Urgency
EduDistilBERT	83.01	89.67	92.43
DistilBERT	82.88	89.12	92.14
EduBERT	82.91	89.78	92.24
BERT-base	82.80	89.47	92.14
Wei et al. (2017)	81.88	86.08	86.68

Future Work & Conclusion. EduBERT and EduDistilBERT are fine-tuned on millions of tokens, in contrast to the billions of tokens required to make the most of the architecture potential (Devlin et al., 2019). We are actively seeking more data to train models even more capable of producing contextualized word representations in the educational domain. We are making EduBERT and EduDistilBERT publicly available in the hope that they will facilitate learning analytics research at large.

REFERENCES

- Agrawal, A., & Paepcke, A. (2014). The Stanford MOOCPosts Dataset. Retrieved from <https://datastage.stanford.edu/StanfordMoocPosts/>
- Alsentzer, E., Murphy, J., Boag, W., Weng, W.-H., Jin, D., Naumann, T., & McDermott, M. (2019). Publicly Available Clinical BERT Embeddings. Proceedings of the 2nd Clinical Natural Language Processing Workshop, 72–78.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proc. Of NAACL.
- Guo, S. X., Sun, X., Wang, S. X., Gao, Y., & Feng, J. (2019). Attention-Based Character-Word Hybrid Neural Networks With Semantic and Structural Information for Identifying of Urgent Posts in MOOC Discussion Forums. IEEE Access, 7, 120522–120532.
- Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. ACL.
- Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2019). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics.
- Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. Proc. of NAACL.
- Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. NeurIPS EMC² Workshop.
- Wei, X., Lin, H., Yang, L., & Yu, Y. (2017). A Convolution-LSTM-Based Deep Neural Network for Cross-Domain MOOC Forum Post Classification. Information, 8(3).
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M. & Brew, J. (2019). HuggingFace's Transformers: State-of-the-art Natural Language Processing. ArXiv, abs/1910.03771.
- Onah, D. F., Sinclair, J., & Boyatt, R. (2014). Dropout rates of massive open online courses: behavioural patterns. EDULEARN14 proceedings, 1, 5825-5834.

Using Eye Tracking Data to Analyze the Effects of Learning Hints in Source Code Comprehension

Fabian Deitelhoff, Andreas Harrer, Benedikt Schröder

University of Applied Sciences and Arts, Dortmund, Germany

{fabian.deitelhoff, andreas.harrer}@fh-dortmund.de

benedikt.schroeder020@stud.fh-dortmund.de

H. Ulrich Hoppe

University of Duisburg-Essen, Germany

hoppe@collide.info

Andrea Kienle

University of Applied Sciences and Arts, Dortmund, Germany

andrea.kienle@fh-dortmund.de

ABSTRACT: Computational thinking has been identified as an essential problem-solving skill in the information age. Although more specialized, programming is an essential manifestation of computational thinking, and in turn, source code comprehension is a vital subskill of programming. The study reported here compares the effects of different types of learning hints supporting source code comprehension. Our analysis relies heavily on using eye tracking data in combination with specific data models and visualizations. This form of behavioral analytics is complemented with answers to comprehension questions to assess the effects of these hints with different code examples. Our findings indicate that syntax highlighting is of limited benefit for better comprehension and a dynamic highlighting of the scope of code blocks and variables is less used than expected. Additionally, we have tried to identify standard code reading patterns using sequence alignment and clustering techniques.

Keywords: learning hints, source code comprehension, eye tracking, visual data analytics, eye movement patterns, AOI-DNA, AOI-STG

1 INTRODUCTION

In a society permeated by digital representations and tools in professional and everyday life, the desirable general knowledge of science, technology, engineering, and mathematics (STEM) must be combined with more meta-level skills like critical thinking, adaptive problem solving, and creativity. "Computational thinking" (CT) is an important ingredient in this context (Wing, 2006). Although CT cannot be reduced to programming, programming is an activity that both builds on CT and can support the development of CT skills. In the area of research on programming and programming education, eye tracking has been used as an analysis method with pioneering work, e.g., by Crosby and Stelovsky (Crosby & Stelovsky, 1990). One advantage of eye tracking is that it is an additional source of objective information that relies on the actual user behavior.

2 APPROACH & HYPOTHESES

The basis of every eye tracking analysis in programming are fixations on specific regions of source code examples. For this purpose, AOIs are placed around every code line (**AOI line model**) or every important workspace area (**AOI block model**). These are marked with letters, to later refer to the specific regions in the source code, so that eye movements can be represented as a string of characters. The fixation calculation is done with an **I-VT** filter. For analyzing the reading behavior of participants, we use a top-down approach with predefined patterns. Two global patterns are the Linear Scan and Jump Control, also known as **Story Order Reading** (SOR) and **Execution Order Reading** (EOR) (Busjahn et al., 2015). To analyze similar reading patterns of participants, we used sequence analysis based on the similarity score, calculated with the Needleman-Wunsch (N-W) algorithm.

Previous research has investigated the effects of syntax highlighting as a form of learning hint. In our study, we are using three different source code examples (Bubble Sort, Greatest Common Divisor, and an object-oriented Vehicle class), with different types of learning hints. We hypothesize that the answer quality varies between these conditions due to a varying support level of available learning hints, and that the learning hints are used differently (also in terms of reading patterns) for the different source code examples.

3 STUDY PROTOTYPE & DATA BASIS

We asked the following comprehension questions for the three different code examples: (**Bubble**) "What does the list look like after two runs of the outer loop?", (**GCD**) "To which values are the variables 'number1' and 'number2' set after three runs of the loop?", and (**Vehicle**) "To which values are the objects 'vOne' and 'vTwo' set at the end of the program?" The three source code stimuli and three learning hint conditions lead to six variants, with the advantages, that (a) the fixed order of code examples reduces the overall complexity of the study, and (b) that there are always two groups with the same condition we can analyze as one group.

We recorded **n = 24** participants from the nearby University campus, out of which seven were females and **17** males, with a mean age of **26.29** (SD = 4.28). The participants were all students between the semesters 1-10, with various Computer Science backgrounds. No participant had to be excluded due to a lack of English reading or programming skills. The eye tracking data quality of all 24 participants was very good to good.

4 ANALYSIS RESULTS

We analyzed the three hint conditions (1) Syntax Highlighting, (2) Dynamic, and (3) Plain (no learning hint active) across the three source code examples. For the answer quality hypothesis, the data shows that (1) is balanced for the correct/incorrect answers. (1) seems not to be an essential factor related to answering a comprehension question for our code examples. The navigational aspect of the underlying source code comprehension may be different, but the core of the source code examples is the same: A complex code is still complicated. However, the difference between (2) and (3) are indecisive for the Bubble and GCD code examples. For the Vehicle code, both conditions with the dynamic help and the plain text seems to have no positive effect on the comprehension result.

To answer the dynamic hint usage hypothesis the dynamic learning hint was not used that often compared to our preliminary assumption. The learning hint seems to be helpful for the GCD code example, but indecisive for the Bubble and not helpful for the Vehicle code examples. This effect can be explained with the overall differences in the difficulty level, the way the code examples were presented, or the prior knowledge of the participants.

To analyze the reading behavior and answer the corresponding hypothesis, we clustered participants with similar eye movement data. This analysis revealed similarities of participants between different source code examples. We could see that participants with a similar reading behavior on one code example tend to have similar reading patterns for other examples, which is an important observation. This result implies that the comprehension strategies of the participants are, to some extent, robust, what is an indicator for comprehension knowledge, independent of the concrete code.

However, the data showed that many participants have problems with the object-oriented code, no matter which learning hint was available. Syntax highlighting could change the way participants perceive source code, but in the core, regarding the complexity, stays the same. This result needs to be analyzed furthermore, to distinguish between eye movement on source code with syntax highlighting and without and the perception of the task difficulty.

5 SUMMARY & DISCUSSION

To our surprise, the Vehicle code example turned out to be the most difficult one in terms of answer correctness. We initially assumed that the Bubble code was the most complex one, due to the complexity of the control structure (nested loops). However, the study showed that many participants had problems with the object-oriented code, no matter which learning hint was available. Syntax highlighting could change the way participants perceive source code, but in the core, like the complexity, stays the same. The dynamic learning hint was less used than expected. We thought that our target group, with knowledge in programming, would use this hint more frequently. Overall, the hint may be useful for the GCD example but ambiguous for the other two code examples. Regarding the reading patterns on both used AOI models, we found common patterns across all participants, code examples, and learning hints. The analysis showed that these patterns form coherent groups through calculating the edit distances (N-W) and with hierarchical clustering. Furthermore, a first analysis showed that these patterns are not the distinguishing factor for the answer quality of participants. The reasons for this finding need to be further analyzed to exclude factors of the study conditions and our participant group.

REFERENCES

- Wing, J. M. (2006). Viewpoint: Computational Thinking. *Communications of the ACM*, 49(3), 33–35. <https://doi.org/10.1145/1118178.1118215>
- Crosby, M. E., & Stelovsky, J. (1990). How Do We Read Algorithms? - A Case Study. *Computer*, 23(1), 25–35. <https://doi.org/10.1109/2.48797>
- Busjahn, T., Bednarik, R., Begel, A., Crosby, M., Paterson, J. H., Schulte, C., ... Tamm, S. (2015). Eye Movements in Code Reading: Relaxing the Linear Order. *IEEE International Conference on Program Comprehension*, 2015-Augus, 255–265. <https://doi.org/10.1109/ICPC.2015.36>

Semantically Adjusting Word Frequency for Estimating Word Difficulty from Unbalanced Corpora

Yo Ehara

Shizuoka Institute of Science and Technology
ehara.yo@sist.ac.jp

ABSTRACT: For second language education, analyzing the words known by learners is crucial. We cannot test all the words in the language vocabulary knowledge of each learner before obtaining it, as second language vocabularies are large. Hence, it is preferable to estimate the difficulty of words for learners by using only the results of tests of their small vocabulary. To this end, the word frequency in a balanced corpus, whose frequencies are not biased towards some domains, has been reported to work well. However, manually balancing the corpus requires significant time and cost. This paper proposes a novel word frequency counting method that automatically adjusts the word frequency of an unbalanced corpus based on the learners' vocabulary test results. The experimental results reveal that the frequency counted by our method can predict learners' vocabulary test responses better than raw frequencies.

Keywords: Second Language Learning, Balanced Corpora, Visualization

1 INTRODUCTION

When learning second languages, learners must acquire many words; hence, estimating the words that each learner currently knows is important for measuring their learning status. By continuously monitoring their status, large amounts of data will be available from each learner (Flanagan & Ogata, 2018). However, if this is not possible, the data must be made easily obtainable, e.g., from vocabulary test results for each learner (Ehara, 2018; Yeung & Lee, 2018). To create such vocabulary tests and interpret their results, the difficulty of each word must be determined. As learners can have different backgrounds (purposes, native languages, and length of learning), automatically estimating the difficulty of words from the vocabulary test result data is desirable. However, as the number of words that can be tested for a learner is limited – typically, up to several hundred – we must estimate the difficulty of the remaining words from the vocabulary test results (Beglar, 2010; Nation, 2006).

To estimate word difficulty, several balanced corpora, such as the British National Corpus (BNC), have been used, as a good correlation exists between their frequency and word difficulty (Chen & Meurers, 2018; Tamayo, 1987). A balanced corpus is one that is manually adjusted to be unbiased towards specific domains, such as politics, sports, or writing styles. However, manual balancing is costly, and the approach for balancing differs among corpora. This is especially problematic when generalizing a successful methodology to make it successful for another language because, in the latter language, we usually cannot obtain a similarly balanced successful corpus. Hence, automatic adjustment of the word frequency of an unbalanced corpus would promote its wide applicability.

This paper proposes such a method. Our method automatically adjusts the word frequency of an unbalanced corpus to make the word difficulty based on the adjusted word frequency fit the

vocabulary test result. Our key idea is to leverage a recent natural language processing (NLP) technique called contextualized word embeddings, from which we can obtain semantic vector representations of each occurrence of the word. Our method can identify and count the occurrences to effectively estimate the difficulty of the word as the adjusted frequency of the word. The experimental results indicate that the adjusted frequencies of the words correlate better with the estimated word difficulty than with the raw frequencies.

2 PROPOSED METHOD

We consider J learners $\{l_1, \dots, l_J\}$ and I words $\{v_1, \dots, v_I\}$. We write the vocabulary test result data as $\{y_{ij}\}$. $y_{ij} = 1$ if learner l_j answered correctly for word v_i and $y_{ij} = 0$ otherwise. Our objective is to estimate the difficulty of each word v_i such that it fits data $\{y_{ij}\}$. As a difficult word can be answered correctly by a skilled learner, we must account for each learner's ability as well. This leads to the following formula, which is also the basis of item response theory (IRT) models (Baker, 2004). Here, a_j is the ability of learner l_j and d_i is the difficulty of word v_i . The logistic sigmoid function $\sigma(x) := (1 + \exp(-x))^{-1}$ is used to normalize $a_j - d_i$ into the range $[0,1]$ to make the resulting value interpretable as a probability.

$$P(y_{i,j} = 1 | l_j, v_i) = \sigma(a_j - d_i)$$

As this formula solely relies on the vocabulary test result data, it cannot estimate the difficulty of words not included in the tests. To achieve this, previous studies (Ehara, 2018; Yeung & Lee, 2018) used K_i , or the raw word frequency of word v_i , as follows: $d_i = -\log(K_i + 1) + D$, where D is a constant to make d_i positive. As log is a monotonously increasing function, the more frequent word v_i becomes, the less difficult d_i becomes. Similarly, in our proposed method, $d_i = -\log(K_i^{adj} + 1) + D_{adj}$, where K_i^{adj} is the adjusted frequency of word v_i , and D_{adj} is a constant.

Here, we define K_i^{adj} as follows. By using a contextualized word embedding method, we can obtain the semantic vector representation \vec{x}_k for each of the K_i occurrences of word v_i in the corpus. Intuitively, the occurrences that are too distant from typical usage are likely to be exceptional usages that the learner may not know even if he/she knows the typical usage of the word. Hence, by counting only the occurrences whose vectors are close to the vector of the typical usage, we can adjust the frequency to exclude such exceptional usages from counting. Let $X_i = \{\vec{x}_1, \dots, \vec{x}_{K_i}\}$ be the set of the vectors for the K_i occurrences of word v_i . Let A be a $T_2 \times T_1$ matrix, and $AX_i = \{A\vec{x}_1, \dots, A\vec{x}_{K_i}\}$. When $T_2 = 2$, A is a projection matrix that projects the vectors into a two-dimensional visualization space. Let $N(A\vec{c}_i, r, AX_i)$ be the number of words within distance r measured from \vec{c}_i . Our model adjusts the word frequency as $K_i^{adj} = N(A\vec{c}_i, r, AX_i)$ by tuning projection matrix A , center point \vec{c}_i of each word v_i , and distance to exclude exceptional usages r .

3 EXPERIMENTS AND CONCLUSIONS

Figure 1 depicts the obtained visualization for the word “period”, where each triangle denotes an occurrence and the red dotted line is the curve denoting radius r . The occurrences within the curve in the right hand are the occurrences that may be known by the learners. Note that this is not merely a visualization; parameters such as transformation matrix A , radius r , and the center point of each word \vec{c}_i are trained and tuned to fit the vocabulary test result dataset.

Figure 2 depicts the adjusted frequency against the raw frequency. The raw frequency was clearly adjusted by removing exceptional usages. Next, the accuracies of predicting the words known to learners using the raw and adjusted frequencies were compared. Although our goal is to adjust the frequency of an unbalanced corpus, comparisons of different corpora may be affected by other factors such as stemming and sentence splitting. To mitigate this, we extracted texts in the *art* domain on BNC, a balanced corpus, and used it as an unbalanced corpus. We used BERT (Devlin, Chang, Lee, & Toutanova, 2019) as the contextualized word embedding method and the publicly available dataset by (Ehara, 2018) to evaluate our model. The accuracy when using the raw word frequency of the *art* domain was 0.61. When using our proposed adjustments to the word frequency by counting the words only within this radius, the accuracy was 0.64. Hence, the adjustment to the word frequency for unbalanced corpus was effective. Our method was also effective when using frequencies of all domains: 0.67 was improved to 0.72. In future work, we will make a visualization similar to that in Figure 1 interactive to help learners learn unfamiliar word usages.

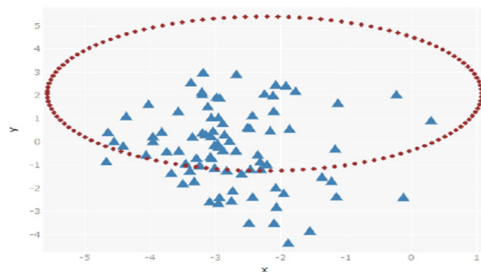


Figure 1: Obtained Visualization

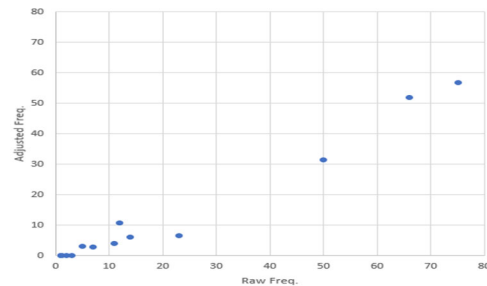


Figure 2: Raw Freq. vs. Adjusted Freq.

ACKNOWLEDGEMENTS

This work was supported by JST, ACT-I Grant Number JPMJPR18U8, Japan. We used the ABCI infrastructure made by AIST for computational resources.

REFERENCES

- Baker, F. B. (2004). *Item Response Theory: Parameter Estimation Techniques, Second Edition*. CRC Press.
- Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing*, 27(1), 101–118.
- Chen, X., & Meurers, D. (2018). Word frequency and readability: Predicting the text-level readability with a lexical-level attribute. *Journal of Research in Reading*, 41(3), 486–510.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proc. of NAACL*, 4171–4186. Minneapolis, Minnesota.
- Ehara, Y. (2018). Building an English Vocabulary Knowledge Dataset of Japanese English-as-a-Second-Language Learners Using Crowdsourcing. *Proc. of LREC*.
- Flanagan, B., & Ogata, H. (2018). Learning Analytics Infrastructure for Seamless Learning. *Proc. of LAK*.
- Nation, I. (2006). How Large a Vocabulary is Needed For Reading and Listening? *Canadian Modern Language Review*, 63(1), 59–82.
- Tamayo, J. M. (1987). Frequency of Use as a Measure of Word Difficulty in Bilingual Vocabulary Test Construction and Translation. *Educational and Psychological Measurement*, 47(4), 893–902.
- Yeung, C. Y., & Lee, J. (2018). Personalized Text Retrieval for Learners of Chinese as a Foreign Language. *Proc. of COLING*, 3448–3455.

Adaptive Learning Guidance System (ALGS)

Ghada El-Hadad, Doaa Shawky and Ashraf Badawi

Center of Learning Technologies, Zewail City of Science and Technology
Gfawzy, dshawky and abadawi@zewailcity.edu.eg

ABSTRACT: This poster presents the conceptual framework of the Adaptive Learning Guidance System ALGS. The system aims to propose a model for adaptive learning environments where two major concerns arising from past studies are being addressed; the marginal role of the teacher, and the need for a big data approach. Most past studies marginalized the teacher's role in adaptive learning system, particularly the online ones. The most notable quality about ALGS is empowering the teacher with the capability of having input in all stages. This is where the hybrid recommendation system plays a crucial role in the 3-stage ALGS architecture. The second issue addressed is the need for big data to enhance the system functionality. The more the data collected by the system, the more efficient its adaptation functionality which makes it difficult for a first-time-run system and/or a first-time user. Accordingly, collaborative filtering is used at first until adequate data about the user interaction are collected. ALGS architecture consists of a user, content, and 3-stage adaptation models.

Keywords: adaptive learning, tutoring system, machine learning, online learning environment, adaptation model.

1 ALGS ARCHITECTURE

The conceptual architecture of ALGS is based on: User, Content, and a 3-stage adaptation models. The user model represents the data about the learner stored in the system (what the system adapts to) (Aleven, McLaughlin, Glenn, & Koedinger, 2016). The content model is the course content offered by the system and its hierarchical structure and logical order (what the system adapts) (Aleven et al., 2016). The adaptation model sets the adaptation strategies and rules (how the system adapts) (Fakeeh, 2017). The adaptation model consists of 3 stages. The first stage is the collaborative filtering. The second stage is the hybrid recommendation system where machine and teacher recommendations are generated. The third stage is the personalization engine. The data coming from all 3 models are combined to optimize suggestions and predictions to both the learner and the teacher simultaneously. The Adaptive Learning Guidance System architecture is illustrated in Figure 1.

2 ALGS ARCHITECTURE COMPONENTS

2.1 User Model

Data about the user is stored to be retrieved later by the system to enhance the personalization procedure (Rodríguez, Ovalle, & Duque, 2015). User attributes are Cognitive traits, learning preferences (Aleven et al., 2016), Knowledge level, Personal data (Rodríguez et al., 2015) as illustrated in Figure 2.

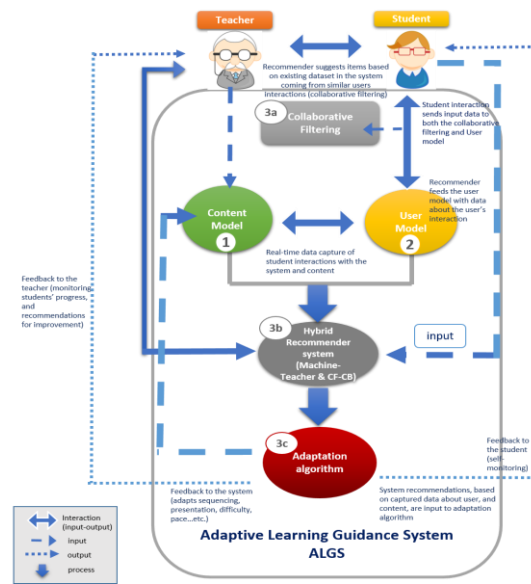


Figure 1: Adaptive Learning Guidance System Architecture

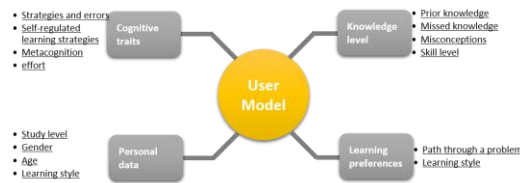


Figure 2: User Model (Learner Attributes)

2.2 Content Model

The model stores data about the knowledge to be presented to the learners and how it is organized in an adaptive manner to achieve the educational goals (Esichaikul, Lamnoi, & Bechter, 2011). The content model consists of content (Rodríguez et al., 2015), and delivery (Esichaikul et al., 2011).

3 ALGS ADAPTATION

3.1 Stage 1: Collaborative Filtering (CF)

When the hypothesized ALGS first runs, the user registers a new user profile. Collaborative filtering (CF) compares these data to a previously installed dataset of similar users and generates related suggestions for the user (Rodríguez et al., 2015). The dual function of CF is eliminating cold starts for first-time users on the one hand (Rodríguez et al., 2015), and allowing for more data gathering about users hence better adaptation on the other (Murray & Perez, 2015).

3.2 Stage 2: Hybrid Recommendation System

The hybrid recommendation system is hybrid on two levels: collaborative filtering (CF)-content-based (CB) hybrid recommender, and machine-teacher hybrid. The combined CF and CB

recommenders together generate more effective predictions (Rodríguez et al., 2015). The result recommendations are based on user's interaction with the system, user's attributes, and data about similar users (Rodríguez et al., 2015). The recommendations address skills that the learner is yet to improve or likely to be weak at for further practice (Drachsler, Hummel, & Koper, 2007). In the machine-teacher hybridization, the generated recommendations are then delivered to the teacher to make decisions. The aroused issue with this step is the level of confidence in the recommendation provided by the system. In other words, the system may provide suggestions that the teacher does not understand the rationale behind in order to make an informed decision (Baker, 2019). Consequently, the recommender sends the teacher a combination of the user's input with the AI analytics. The teacher then decides how the recommendations are handled.



Figure 3: Hybrid Recommendation System

3.3 Stage 3: Personalization Engine

The user attributes, content attributes, along with the 3 stages of adaptation are then integrated together to formulate efficient recommendations for both the student and the teacher about what material needs to be studied next (Fakeeh, 2017). The teacher has a crucial role at directing this stage being the decision maker of selecting, adding, or removing the next adaptation procedure.

REFERENCES

- Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2016). Instruction based on adaptive learning technologies. *Handbook of research on learning and instruction*. Routledge.
- Baker, R. S. (2019). Challenges for the Future of Educational Data Mining: The Baker Learning Analytics Prizes. *JEDM| Journal of Educational Data Mining*, 11(1), 1-17.
- Drachsler, H., Hummel, H., & Koper, R. (2007). Recommendations for learners are different: Applying memory-based recommender system techniques to lifelong learning.
- Esichaikul, V., Lamnoi, S., & Bechter, C. (2011). Student modelling in adaptive e-learning systems. *Knowledge Management & E-Learning: An International Journal*, 3(3), 342-355.
- Fakeeh, K. A. Developing Virtual Class Room Models with Bio Inspired Algorithms for E-Learning: A Survey for Higher Technical Education for Saudi Arabia Vision 2030.
- Murray, M. C., & Pérez, J. (2015). Informing and Performing: A Study Comparing Adaptive Learning to Traditional Learning. *Informing Science: The International Journal of an Emerging Transdiscipline*, 18, 111.
- Rodríguez, P. A., Ovalle, D. A., & Duque, N. D. (2015, August). A student-centered hybrid recommender system to provide relevant learning objects from repositories. In *International Conference on Learning and Collaboration Technologies* (pp. 291-300). Springer, Cham.

Teachers' Sense-Making of Learning Analytics: what is missing?

Diana Forero Tobon, Canan Blake, Yvonne Vezzoli, Mina Vasalou and Manolis Mavrikis

UCL Knowledge Lab, UCL Institute of Education, London, UK

{diana.tobon.18; canan.blake; y.vezzoli; a.vasalou; m.mavrikis}@ucl.ac.uk

ABSTRACT: This study aims at broadening our understanding of teachers' sense-making process of learning analytics visualisations. Using sense-making theory, we analysed six teachers' diaries and semi-structured interviews of the processes they follow in interpreting dashboard visualisations in two different platforms. The results indicate that the sense-making process is highly influenced by teachers' objective and their knowledge of the students. Moreover, it was found that expectations and attributions played a major role during the reading and interpretation processes before resulting in a formulation of a plan. The findings informed two principles for future designs: filter the data according to the teacher's objective, and compare expectations with the data retrieved.

Keywords: Data visualisation, learning analytics, sense-making.

1 INTRODUCTION

Teachers have traditionally played an essential role in the collection, selection and use of data about their students' learning. With the advent of dashboards (Schwendimann et al. 2019), it is important to understand more about the process by which teachers interact and make sense of them. Such understanding would facilitate the design of dashboards that empower teachers during the digital data interpretation process, without intervening on the interpretive process (Duval, 2011). Understanding the teacher's creation of meaning could help dashboard designers take the right choices e.g. choose what to make visible and what to leave behind (Dillenbourg et al., 2011).

Taking a sense-making lens, this study seeks to show how teachers work with data visualizations in dashboards, and to analyse the sense-making process that teachers follow to interpret them. Our aim is to identify design principles that could support future dashboard design. We follow Verbert et al. (2013) who describe the sense-making process in four stages: *awareness* (drawing attention to data), *reflection* (asking about the relevance of data), *sense-making* (creating new insights, answering questions), and *impact* (creating new meaning and change). Using mixed methods, we collected teacher diaries and subsequently held semi-structured interviews with six school and university teachers using two different dashboards: Edmodo and Cambridge Learning Management System (CLMS). We analysed data following Charmaz's (2006) coding process. An iterative approach was employed to explore and compare sentence by sentence of each interview. Each iteration included memos with the relationship between our data and emergent concepts representing the sense-making process and resulting in Figure 1.

2 THE SENSE-MAKING PROCESS AND DESIGN IMPLICATIONS

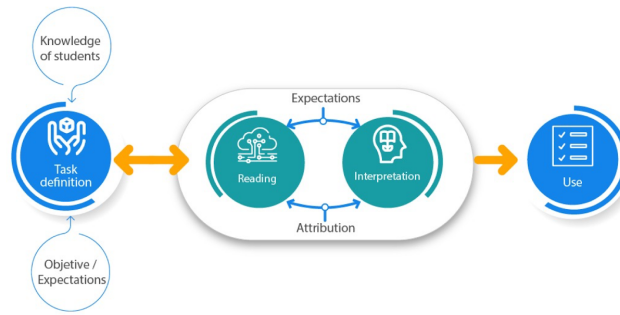


Figure 1: Sense-making process of digital data visualisation

Our findings indicated that, as shown in Figure 1, **task definition** is the first step of the teachers' sense-making process. In this step, teachers first consider what they know about their students and select the tools to check relevant data using the dashboard (e.g. measuring individual progress). This is also informed by the teacher's perceptions and experiences with technology, as well as students' realities and needs. The objective when approaching the platform also informs this step and changes depending on the classroom setting by taking into account beliefs and perceptions of technology.

The goals shared by the participants included checking homework for assessment or class planning purposes, finding data such as students' progress, content coverage and task completion. In each case, these goals filtered the dashboard data that teachers focused on. A design implication is to provide teachers with the option to filter the data available according to their goals. The teacher would be able to find a list of objectives such as "check homework" and "check registration" among others. A goal-oriented approach to identifying relevant data could shorten the time and simplify the sense-making process, as teachers would not need to find the relevant data among all the analytics provided by the platform. However, careful research would be needed to understand the different objectives that teachers may have and the most relevant data that they use to achieve them.

During the second step in Figure 1, **reading and interpretation**, teachers consider the visual elements in the data visualisation such as size and colours and rely on these to detect patterns, problems or abnormalities in the data. Data reading and data interpretation are iterative processes and are supported by sourcing additional data provided by the platform, which teachers consult only when there is an abnormality. In those cases, data is evaluated by comparing it with what teachers have seen in previous, similar situations in the classroom. Data is then validated by comparing it with what teachers attribute the results to and the purpose that teachers have for it in the classroom.

Teachers' constant comparison of data with expectations reveals their need to summarise it and compare it to create an argument, but the fact that teachers were not often explicitly aware of their expectations makes the interpretation process complex. For this reason, designing a tool that helps teachers identify their expectations would support their data reading process, which could also result in simplifying the sense-making process. This would also give teachers the possibility to include the expected outcome of their objective in their analysis. Subsequently, the platform could provide a comparison of the data that is expected and the data that was achieved by the class to support the interpretation process without completing it for teachers (Duval, 2011).

The final stage in Figure 1 is **use of data**. In this stage, teachers used what they understood from the data to make an action plan, which in this study included cases such as following-up on students' progress, evaluating their own teaching practices, and finding opportunities for class improvement. At this stage, designers could provide teachers with tools to extract and/or report specific data summaries tailored to the teacher's action plans for the class.

3 CONCLUSION

This study used sense-making theory to understand how dashboard designs could help teachers make sense of learning dashboards. Although not all dashboards are the same (Schwendimann et al. 2019), it is important to work towards a better understanding of a general sense-making process in this context to identify how design could better support it. Our study made clear that teachers draw upon prior knowledge and other information about their students to decide how they will approach the tasks and goals that they are undertaking in the dashboard. Typically though, dashboards are not tailored according to individual teachers' needs (c.f. Rodriguez-Triana et al., 2018; Schwendimann et al., 2019). Considering the complexity of some of these dashboards, this lack of goal-orientation, despite being a principle of good design, may be leading to misinterpretations of the students' learning process. Similarly, there seems to be a need for supporting data interpretation based on what a teacher may already expect to see in the data. Such a feature has the potential to challenge or verify their hypothesis or beliefs. Emerging research outside education has recently proposed frameworks for integrating users' beliefs in designing and supporting interaction with data (Nguyen et al., 2019). Our study proposes to take this into account in future designs of tools for teachers.

REFERENCES

- Borghouts, J., Soboczenski, F., Cairns, P., Brumby, D.P., (2015). Visualizing Magnitude: Graphical Number Representations Help users Detect Large Number Entry Errors. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, 59, 591-595.
- Charmaz, K. (2006). *Constructing grounded theory: a practical guide through qualitative analysis*. London: SAGE Publications
- Dillenbourg, P., Zufferey, G., Alavi, H., Jermann, P., Do-Lenh, S., Bonnard, Q., et al. (2011). Classroom orchestration: the third circle of usability. In *Proceedings of the 9th international conference on computer supported collaborative learning*, Hong Kong, China, Vol. 1, (pp. 510–517).
- Duval, E., (2011) Attention please!: Learning Analytics for Visualization and Recommendation. In *Proceedings of LAK '11. ACM*, New York, USA, pp.9-17.
- Nguyen, F., Shrestha, S., Germuska, J., Kim, Y. S., & Hullman, J. (2019). Belief-Driven Data Journalism. *Computation + Journalism Symposium 2019*. University of Miami, FL, USA 1-2 February 2019.
- Rodríguez-Triana, M. J., Prieto, L. P., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2018). *The teacher in the loop: Customizing multimodal learning analytics for blended learning* (pp. 417–426). ACM Press <https://doi.org/10.1145/3170358.3170364>.
- Schwendimann, B. Rodríguez-Triana, M.J., Vozniuk, A., Prieto, L.P., Boroujeni, M., Holzer, A., Gillet, D., Dillenbourg, P. (2017). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Trans. Learning Technol.*, vol. 10, no. 1, pp. 30-41.
- Verbert, K., Duval, E., Klerkx, J., Govaerts, S., Santos, J.L. (2013) Learning Analytics Dashboard Applications. *Am. Behav. Sci.* 57, 1500-1509

A proactive perspective on the future of Learning Analytics: A systematic literature review

Stephanie Gaaw

TU Dresden, Center for Quality Analysis
stephanie.gaaw@tu-dresden.de

Cathleen M. Stuetzer

TU Dresden, Center for Quality Analysis
cathleen.stuetzer@tu-dresden.de

ABSTRACT: After several years of research on potential benefits of Learning Analytics (LA) and various approaches to adopt them, users and developers are confronted with a broad mixture of different forms of LA dealing with various challenges, conditions and goals, that may be relevant to their implementation. It is therefore a challenge in itself to navigate through this broad field of research and gain an overview of problems that may arise in this context. In order to face this issue, this contribution (poster) seeks to answer the question, which challenges regarding LA are currently highlighted and need to be addressed over the next five to ten years. Hence, we conducted a systematic literature review to provide a purposeful synthesis on the research field of LA from 2017 up to now. For the exploration of the literature an analytical approach based on a multi-perspectival coding concept was applied. Main results of this contribution are the identification of current challenges in LA and thus a comprehensible overview of how the future of the field can be shaped to enhance the possibility of successful adoption processes, especially in higher education.

Keywords: Learning Analytics, Academic Analytics, Higher Education, Trends and Challenges

1 INTRODUCTION

In the last decades Learning Analytics (LA) has risen to a cutting edge discipline. Through the relation to various disciplines a vast number of modern research techniques and methods are applied. Although this multidisciplinary position promises potential to deepen the understanding and optimization of learning processes and environments, current LA approaches do not accomplish the overall objective to make a “purposeful shift to move from exploratory models to more holistic and integrative systems-level research” (Dawson, Joksimovic, Poquet and Siemens, 2019, p. 446). After several years of research on LA, researchers and practitioners are still facing serious defiances regarding this. Quite contrary, LA research is still all about understanding potentials and challenges in the adoption of LA and discussing the outcomes of a series of attempts to implement them for different purposes (El Alfy, Marx Gómez & Dani, 2018). Particularly in the field of higher education, this leads to a mixture of many different approaches of LA dealing with various challenges, general conditions and goals, that may be relevant to the implementation of LA (e.g. Chen & Zhu, 2019). Therefore, it doesn't come as a surprise that there is already a large number of literature reviews regarding this topic. However, due to the rapid pace of technological progress new work and emphases are constantly being added at high speed. Hence, for outlining related work on current challenges of LA reviews published from 2017 to 2019 were taken into account. With respect to limitations in these studies this review focuses on a broad overview without limitation to a particular technique of LA or a specific sub-area in which LA is used.

2 METHOD

With regards to our goals of this paper we conducted a systematic literature review with focus on the following research question: What are necessary fields of action and challenges of LA which are currently highlighted and need to be addressed in the next five to ten years? Considering that LA is as interdisciplinary as the research field itself, this review was conducted by using international databases, namely Web of Science, Science Direct, Academic Source Complete and Business Source Complete, which are backed up by a multidisciplinary literature pool. Due to the rapid pace of progress in LA and to follow on from the results of the aforementioned reviews, the search period was set from 2017 up to now. Language was limited on German and English. Regarding the search terms, following words were included: learning analytics, challenge, benefit, potential, problem, gap, possibility, need to, future, further research. The initial search in the databases resulted in 220 papers. After selecting relevant papers by applying various selection criteria and removing double entries, the final data set contained 78 papers, which were analyzed in terms of type of paper, used research approaches, discussed challenges and necessary fields of action for further adoption of LA and research. To give an orientation throughout the various challenges we characterized them based on the following *action dimensions (AD)*: *actor-oriented (A)*, *technological (T)*, *methodical (M)*, *organizational (O)*, *ethical/legal (E/L)* and *socio-cultural (SC)* (Gaaw & Stützer, 2017). The papers were coded independently by the authors and the final coding was reviewed together afterwards.

3 RESULTS

By conducting the analysis, it was found that in the last three years methodological, actor-related and technological challenges and fields of action were main topics of discussion for LA researchers. Contributions that provide a socio-cultural perspective on the field are least represented. Overall, it can also be stated that respective challenges are only rarely addressed as individual fields of action. Most of the studies examined discussed at least two dimensions with a view to current or upcoming challenges. With regard to these different perspectives, methodological challenges show a close link to technological circumstances. Furthermore, it can be observed that researchers focus on various approaches, such as visual and predictive analytics, and also discuss benefits of further analysis options, such as social network analysis, text mining or machine learning. From a technological point of view, this is accompanied by discussions on the use of computational techniques, sensor data, wearables and cloud computing. With regard to resulting multimodal data, there is also the challenge of interoperability. Regarding actor-related challenges, researchers are primarily concerned with didactical fields of action that relate to specific educational methods, such as self-regulated learning, motivation of students and personalized learning. Both from an ethical and organizational perspective, which frequently go hand in hand, data protection in general as well as a critical awareness regarding the handling of data generated and exploited through LA is widely discussed. At the organizational level researchers are concerned with the responsibility of institutions for the sensitive handling of data, both with regard to the privacy of learners and the further use of data for quality assurance in teaching. In addition to legal issues, a socio-cultural perspective is also occasionally used, which, for example, discusses the necessity for so-called critical LA in order to reflect on the potential influence of LA on power structures in learning processes or negative effects of this. Another important finding is that the challenges highlighted can no longer be viewed in isolation, but must increasingly be focused on in terms of their reciprocal

interdependencies. As shown in Table 1 various contributions do approach this perspective by simultaneously contemplating several dimensions of action and further challenges and trends.

Table 1: Further challenges & trends – An exemplary overview

AD	Further challenges & trends	Examples
T-M	data analytic techniques: machine learning, computational techniques, new_multimodal_contextual_data sources, predictive analytics, data management, educational theory/design	Hlostá, M., Zdrahal, Z., & Zendaika, J. (2018). Are we meeting a deadline? Classification goal achievement in time in the presence of imbalanced data. <i>Knowledge-Based Systems</i> , 160, 278–295.
A-O	data analytic techniques: corporate learning analytics, institutional responsibility/concerns	Ho, M. (2019). What Have We Learned? TD: Talent Development, 73(1), 12–12.
A-M	educational theory/design, learning experience, applications & tools, data analytic techniques, new_multimodal_contextual_data sources, data management	Chen, B., Chang, Y.-H., Ouyang, F., & Zhou, W. (2018). Fostering student engagement in online discussion through social learning analytics. <i>Internet & Higher Education</i> , 37, 21–30.
A-E/L	data/privacy protection, data ethics, applications & tools, quality assurance, impact evaluation, data management	Watson, C., Wilson, A., Drew, V., & Thompson, T. L. (2017). Small data, online learning and assessment practices in higher education: A case study of failure? <i>Assessment & Evaluation in Higher Education</i> , 42(7), 1030–1045.
A-T	educational theory/design, applications & tools	Conde, M. A., Garcia-Penalvo, F. J., Fidalgo-Blanco, A., & Luisa Sein-Echaluce, M. (2017). Can We Apply Learning Analytics Tools in Challenge Based Learning Contexts? In Zaphiris, P and Ioannou, A (Hrsg.), <i>LEARNING AND COLLABORATION TECHNOLOGIES: TECHNOLOGY IN EDUCATION</i> , PT II (S. 242–256).
A-T-M	educational theory/design, learning experience, data analytic techniques	Dessi, D., Fenu, G., Marras, M., & Reforgiato Recupero, D. (2019). Bridging learning analytics and Cognitive Computing for Big Data classification in micro-learning video collections. <i>Computers in Human Behavior</i> , 92, 468–477.
T-M-O	data analytic techniques, predictive analytics, institutional responsibility/concerns	Kaya, I. E. (2019). Artificial Neural Networks as a Decision Support Tool in Curriculum Development. <i>International Journal on Artificial Intelligence Tools</i> , 28(4), N.PAG-N.PAG.
A-M-O	institutional responsibility/concerns, validity_quality of LA	Ilfenthaler, D. (2017). Are Higher Education Institutions Prepared for Learning Analytics? <i>TechTrends: Linking Research & Practice to Improve Learning</i> , 61(4), 366–371.
A-O-E/L	data ethics, quality assurance, institutional responsibility/concerns	Wintrup, J. (2017). Higher Education's Panopticon? <i>Learning Analytics, Ethics and Student Engagement</i> . <i>Higher Education Policy</i> , 30(1), 87–103.
A-T-E/L	data/privacy protection, data ethics	Gursoy, M. E., Inan, A., Nergiz, M. E., & Saygin, Y. (2017). Privacy-Preserving Learning Analytics: Challenges and Techniques. <i>IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES</i> , 10(1), 68–81.
T-M-E/L	data/privacy protection, data analytic techniques	Hoel, T., & Chen, W. (2019). Privacy engineering for learning analytics in a global market: Defining a point of reference. <i>International Journal of Information & Learning Technology</i> , 36(4), 288–298.
A-T-O	applications & tools, quality assurance, institutional responsibility/concerns	Armatas, C., & Spratt, C. F. (2019). Applying learning analytics to program curriculum review. <i>International Journal of Information & Learning Technology</i> , 36(3), 243–253.
M-O-E/L	data/privacy protection, applications & tools, institutional responsibility/concerns, data management, validity_quality of LA	Sherriff, G., Benson, D., & Atwood, G. S. (2019). Practices, Policies, and Problems in the Management of Learning Data: A Survey of Libraries' Use of Digital Learning Objects and the Data They Create. <i>The Journal of Academic Librarianship</i> , 45(2), 102–109.
M-O-E/L-SC	data/privacy protection, data ethics, applications & tools, institutional responsibility/concerns: complexity & interconnectivity; validity_quality of LA	Gasevic, D., Tsai, Y.-S., Dawson, S., & Pardo, A. (2019). How do we start? An approach to learning analytics adoption in higher education. <i>International Journal of Information & Learning Technology</i> , 36(4), 342–353.
A-M-O-E/L	data/privacy protection, data ethics, learning experience, institutional responsibility/concerns	Jones, K. M. L. (2019). „Just Because You Can Doesn't Mean You Should": Practitioner Perceptions of Learning Analytics Ethics. <i>Portal: Libraries & the Academy</i> , 19(3), 407–428.
A-T-M-O	data representation, applications & tools, data analytic techniques, institutional responsibility/concerns, data management	Rienties, B., Cross, S., Marsh, V., & Ullmann, T. (2017). Making sense of learner and learning Big Data: Reviewing five years of Data Wrangling at the Open University UK. <i>Open Learning</i> , 32(3), 279–293.
A-T-M-E/L	data representation, data/privacy protection, data ethics, data analytic techniques, predictive analytics, new_multimodal_contextual_data sources: Industry 4.0, sensor data	Koren, I., & Klamann, R. (2018). Enabling visual community learning analytics with Internet of Things devices. <i>Computers in Human Behavior</i> , 89, 385–394.
A-M-E/L-SC	data management, data/privacy protection, data ethics: disempowerment, blackboxing	Knox, J. (2017). Data Power in Education: Exploring Critical Awareness with the "Learning Analytics Report Card". <i>Television & New Media</i> , 18(8), 734–752.
T-M-O-E/L	data/privacy protection, new_multimodal_contextual_data sources, institutional responsibility/concerns, data management, real-time feedback_evaluation	Estévez-Ayres, I., Arias Fisteus, J., & Delgado-Kloos, C. (2017). Lostregio: A distributed stream-based infrastructure for the real-time gathering and analysis of heterogeneous educational data. <i>Journal of Network and Computer Applications</i> , 100, 56–68.
A-T-O-E/L	data/privacy protection, data ethics, institutional responsibility/concerns, validity_quality of LA	Zilvinskis, J., & Borden, V. M. H. (2017). Concluding Thoughts. <i>New Directions for Higher Education</i> , 2017(179), 103–108.
A-T-M-O-E/L	data/privacy protection, data ethics, data analytic techniques: big data, institutional responsibility/concerns: stakeholder	Johnson, J. A. (2017). Ethics and Justice in Learning Analytics. <i>New Directions for Higher Education</i> , 2017(179), 77–87.
A-T-M-O-E/L-SC	student motivation and engagement, data ethics: power and politics in online learning, institutional responsibility/concerns, validity_quality of LA, data analytic techniques: big data, critical LA	Scott, J., & Nichols, T. P. (2017). Learning analytics as assemblage: Criticality and contingency in online education. <i>Research in Education</i> , 98(1), 83–105.

Based on the results of the conducted literature review, this contribution discussed current challenges and action dimensions in LA research. With regard to Table 1, the interconnectivity between these different dimensions were highlighted. In this way, this contribution also pinpoints trends in the field of LA in order to provide researchers and practitioners with a comprehensible overview on how to shape the future of the field to enhance the possibility of successful implementation and adoption processes is met.

REFERENCES

- Chen, B., & Zhu, H. (2019). Towards value-sensitive learning analytics design. Presented at the ACM International Conference Proceeding Series.
- Dawson, S., Joksimovic, S., Poquet, O., & Siemens, G. (2019). Increasing the impact of learning analytics. Paper presented at the ACM International Conference Proceeding Series.
- El Alfy, S., Gomez, J. M., & Dani, A. (2018). Exploring the benefits and challenges of learning analytics in higher education institutions: a systematic literature review. *Information discovery and delivery*, 47(1), 25–34.
- Gaaw, S., & Stützer, C. M. (2017). Learning und Academic Analytics in Lernmanagementsystemen (LMS): Herausforderungen und Handlungsfelder im nationalen Hochschulkontext. 20. Workshop GeNeMe'17, 145–161.

Using Writing Logs to Help Validate Assessments for Learning

Hongwen Guo, Mo Zhang, Paul Deane, and Randy E. Bennett

Educational Testing Service

hguo@ets.org, mzhang@ets.org, pdeane@ets.org, rbennett@ets.org

ABSTRACT: A theoretically driven approach to assessment of, as, and for learning led to a scenario-based assessment (SBA) design to engage and assist students in writing. The structure of the SBA simulates a condensed writing project undertaken in an order that a skilled practitioner might follow: A scenario (or topical context) is presented along with source materials; students are then given a sequence of lead-in tasks that require reading and summarizing arguments in the sources, critiquing those arguments, analyzing them, and finally composing an argument essay on the same topic by presenting a position with appropriate reasoning using evidence from the sources. In this study, we analyzed students' keystroke logs and investigated how the SBA structure impacted students' writing performance and processes using stochastic process modeling methods. To the extent that this SBA structure is proven effective, it can be used as a model for classroom instruction or be embedded into the curriculum design.

Keywords: Process modeling, keystroke logs, writing states

1 ASSESSMENT FOR LEARNING AND TEACHING

The scenario-based assessment (SBA) design discussed in this study is a theoretically driven approach to engage and assist students in writing assessment (Bennett, et al., 2016; Deane, et al., 2011). The structure of this SBA design simulates a condensed writing project undertaken in an order that a skilled practitioner might follow: A scenario (or topical context) is presented along with source reading materials; students are then given a sequence of lead-in tasks that require reading and summarizing arguments in the sources, critiquing those arguments, analyzing them, and finally composing an argument essay presenting a position and reasoning using evidence from the sources. As a result, rich and potentially diagnostic information about students' performance and skills can be obtained from SBA. Further, such SBA design is theorized to increase students' engagement by providing a reasonably realistic setting on a common topic. The lead-in tasks facilitate students' engagement with the sources, reducing differences in topic familiarity among the writers, while activating various skills needed for completing the culminating essay task.

In this study, we aim to understand what impacts the lead-in task had on students writing processes. An experiment was conducted using the original SBA (as described above) and its variant form (Zhang, et al., 2019). In the variant form, students were asked to write an essay on the same topic without the benefit of lead-in tasks. Data were collected from more than 800 8th-grade students from eight volunteer schools in New Jersey, Delaware, West Virginia, South Carolina, Alabama, Minnesota, South Dakota, and Utah in the US. Essays were scored by trained raters on two rubrics. Rubric 1 focuses on writing fundamentals and conventions, such as grammar, mechanics, word use,

and sentence structure; Rubric 2 evaluates higher-level writing skills specific to argumentation, such as the quality of the reasoning and supporting evidence. Both rubrics have a scale from 1 to 5, although 0 may be given to empty or non-English essays. We excluded essays receiving a 0 score from analysis. Students' writing processes on the essay task were recorded using keystroke logs (Leijten & Van Waes, 2013).

An initial group comparison revealed that students who took the original SBA form (Group 1) and students who take the variant form (Group 2) were comparable on their keyboarding skill (measured based on in-word characters per minute on common English words), and that the essay task position did not impact students' essay scores. The mean total essay score was about 4.5 for both groups. However, compared to Group 2, the mean total essay writing time was shorter (700 seconds vs. 900 seconds) and the mean essay length in terms of number of words was shorter (150 words vs. 200 words) for Group 1.

2 METHODS TO ANALYZE WRITING KEYSTROKES

Kellogg (2001) and Hayes (2012) proposed a theoretical writing model that decomposed the writing process into four states: planning, translation, transcription, and revision. In the *Planning* state, writers conduct task analysis, idea generation and text organization activities. In the *Translation* state, writers' activities mainly include the linguistic operations necessary to express ideas in words and sentences. In the *Transcription* state, writers render that language on paper or on screen. Finally, the *revision* state involves reviewing and amending the text to correct errors or to otherwise improve the text content or the plan underlying it.

Drawing upon the above theory, we classified the sequence of keystrokes/actions into four writing states: E (editing), J (jump editing), P (long pause), and T (text producing) (Guo et al., 2019). Figure 1 shows a sequence of a single student's writing states during composition.

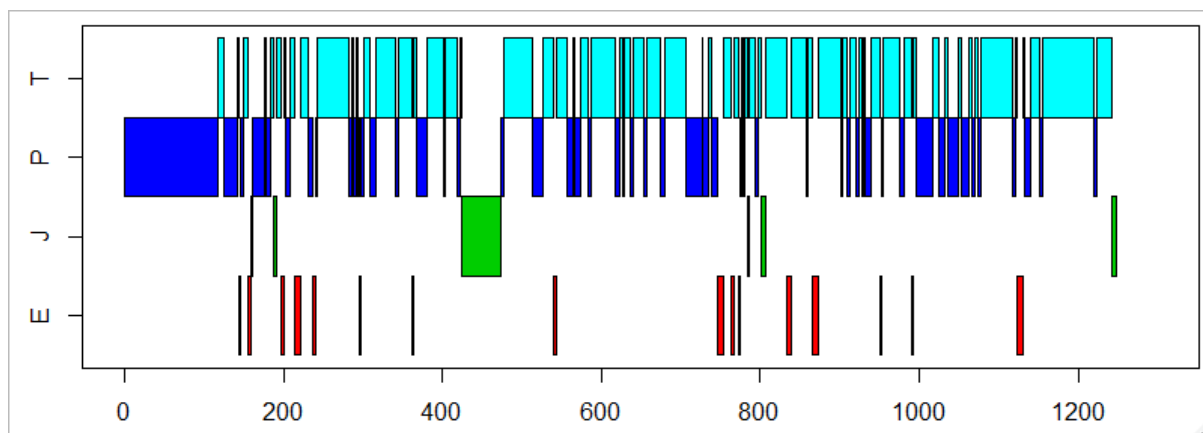


Figure 1: A state sequence of one student's writing process. The x-axis stands for time in seconds, and the y-axis for state of E (Editing), J (Jump Editing), P (Long Pause), and T (Text Production).

Next, students' writing states, their duration in time, and transitions from one state to another were used to investigate the differences in writing process between Group 1 and Group 2. This investigation was aimed at understanding how the difference in essay position affected how students composed their essays. Given that the writing states are discrete, and the duration times

are continuous, Markov-type processes were considered. We compared different Markov models and found that the semi-Markov model (Guo, et. al., 2019; Krol & Saint-Pierre, 2015) had a better fit to the writing-process data.

3 RESULTS AND DISCUSSION

Our results revealed that, as noted, that the order of the tasks did not have significant impact on essay scores. But the placement of lead-in tasks prior to the essay enabled students to produce essays similar in quality in less time using fewer words. We also found that, from the writing keystroke logs and stochastic process modeling, students who took the original scenario-based design (Group 1) had longer durations in the editing and text production states and shorter durations in the long-pause state transiting to text production than students who wrote the essay first without the benefit of the lead-in tasks (Group 2). In addition, Group 1 students had relatively fewer, but longer, jump-editing behaviors, indicating more concentrated editing and perhaps greater efficiency.

Because the SBA consists of a sequence of items that model the skills that students are expected to learn, this structure has practical value for assisting instruction in writing classes. Teachers can embed this structure into their curriculum design. Our results have revealed advantages of SBA in eliciting more efficient writing processes by students. Overall, the original SBA form with the theoretically motivated assessment design may serve to reduce students' working memory load while they plan and review their essays, which would lead to more efficient, fluent text production, freeing student time and cognitive resources to scan the text to monitor and correct problems. It will be of value for future studies to collect data from classrooms that implement such SBA designs and investigate the effect on student learning and writing performance.

REFERENCES

- Bennett, R. E., Deane, P., & van Rijn, P.W. (2016). From cognitive-domain theory to assessment practice. *Educational Psychologist*, 51, 82-107.
<https://doi.org/10.1080/00461520.2016.1141683>
- Deane, P., Fowles, M., Baldwin, D., & Persky, H. (2011). The CBAL summative writing assessment: A draft eighth-grade design (RM-11-01). Princeton, NJ: Educational Testing Service.
- Guo, H., Zhang, M., Deane, P., & Bennett, R. (2019). Writing process differences of subgroups reflected in keystroke logs. *Journal of Educational and Behavioral Statistics*, 44, 571-596.
<https://doi.org/10.3102/1076998619856590>
- Hayes, J. (2012). Modeling and remodeling writing. *Written Communication*, 29, 369-388.
- Kellogg, R. T. (2001). Competition for working memory among writing processes. *The American Journal of Psychology*, 114, 175 – 191.
- Krol, A., & Saint-Pierre, P. (2015). SemiMarkov: An R package for parametric estimation in multi-state semi-Markov models. *Journal of Statistical Software*, 66, 1-16.
- Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research using input log to analyze and visualize writing processes. *Written Communication*, 30, 358-392.
- Zhang, M., van Rijn, P., Deane, P., & Bennett, E. R. (2019). Scenario-based assessments in writing: An experimental study. *Educational Measurement: Issues and Practice*, 24, 73-90.
<https://doi.org/10.1080/10627197.2018.1557515>

Why Predictions of At-Risk Students Are Not 100% Accurate? Showing Patterns in False Positive and False Negative Predictions

Martin Hlosta, Zdenek Zdrahal, Vaclav Bayer, Christothea Herodotou

The Open University
martin.hlosta@open.ac.uk

ABSTRACT: Predictive modelling with the focus on identification of students at risk of failing has become one of the most prevalent topics in the Learning Analytics and Educational Data Mining. Most of the published work is focused on training the machine learning model that achieves the highest prediction performance, as measured by several metrics. Nevertheless, limited work focuses on the behaviour of the model and in particular, analysis of the errors the models make during predictions. This poster presents preliminary results that fill this gap by providing a methodology for finding the patterns of errors both for False Positives and False Negatives. We show results from the task of predicting students at risk of not submitting their first assignments on 48 first-year STEM courses, separately for False Positives and False Negatives. The erroneous predictions that are not possible to be explained will inform subsequent qualitative analysis i.e., interviews with students.

Keywords: Predictive Modelling, At-risk students, Error analysis, Higher Education.

1 INTRODUCTION AND RELATED WORK

Learning Analytics (LA) is a cross-disciplinary field where machine learning models help to understand or even improve student learning. Yet, even the best Machine Learning models will still produce errors. (Kitto, Shum, & Gibson, 2018) argue that striving for best performance metric should not be the goal of LA and having imperfect models does not necessarily mean that they should not be used. Algorithmic error presentation has been recognised as one of the factors to focus on during designing user-facing predictive systems (Springer & Whittaker, 2018). More important, people tend not to trust predictions if they see that the algorithm is making errors (Dietvorst, Simmons, & Massey, 2015). Understanding the errors that the machine learning models tend to make and communicate it efficiently might be a factor that will increase the user acceptance of such systems. Despite the increasingly better performance of the models, research analysing the errors that the models make is very limited.

We identified only two papers directly analysing the error in Predictive LA, in both cases not as the primary contribution. (Lakkaraju et al., 2015) utilised an approach based on Frequent Patterns and Association Rule mining to describe the errors of the classification models. The method first identifies all the frequent patterns covering at least 80% of students. Sorted by false predictions ratio, they show only the top 2 patterns for each algorithm. For example, the most prevalent error is done on students with high GPA and high absence rate. The paper did not distinguish between False Positives (FP) and Negatives (FN). (Qiu et al., 2016) provided as part of the paper for predicting assignment grades and certificate completion an error analysis, splitting errors in (1) unpredictable failing students, (2) unpredictable “succeeders” and (3) swinging cases. For group (2) by conducting

interviews, a large proportion of those students shown to be those who took a similar course offline before.

There is a plethora of research dedicated to predictive analytics in education focusing on improving the accuracy, mostly in MOOCs (Gardner & Brooks, 2018). However, very little work has focused on understanding these models themselves. This work aims to fill this gap. In particular, we aim to analyse whether (1) we can identify patterns in predictions where the models make more errors and (2) what the difference is between these patterns for FP and FN.

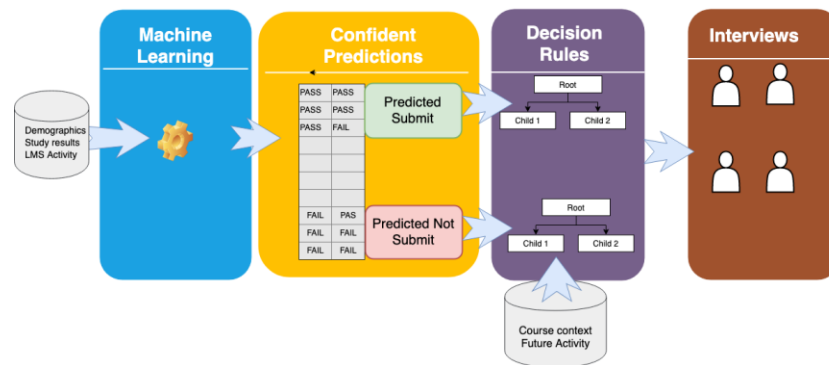


Figure 1: Proposed methodology

2 METHODOLOGY

We computed weekly predictions in 219 undergraduate courses focused on identifying students at risk of not submitting their next assignment, a proven proxy for failing the course (Hlosta, Zdrahal, & Zendulka, 2017). We used the previous run of the same course to compute the predictions and used students' LMS activity, previous results and demographics. We analysed 37,119 predictions in 48 STEM courses between 2017 and 2019 semesters with 25,847 unique students. We narrowed the focus on predicting the first assignment (A1) as most of the students drop out even here and we selected the predictions three weeks before the deadline. We utilise Gradient Boosted Machine (GBM), proven as the best model after several years of tuning, optimised for the highest ROC AUC. The AUC is 85.9, F1=54.1, Precision=61.6, Recall=48.2, considering that NotSubmit is a positive class.

To exclude borderline cases, we only considered confident predictions, which we defined as having the predicted confidence of not submitting greater than 0.8 or lower than 0.2. These give 26,034 predictions in two groups (1) 2,581 True positives (TP), 1,610 FP and (2) 20,326 True Negatives (TN) and 1,517 FN. The features used for training were enhanced by (a) the context of the predictions, i.e. various course information such as average online activity, submission ratio or number of registered students; and by (b) information from the future, i.e. activity of the students after the predictions are generated.

We then trained two decision tree models, one classifying FP and TP and one classifying FN and TN. with the minimum leaf size=30, depth=4 and extracted all the rules for FP and FN respectively. Our proposed methodology is depicted in Figure 1.

3 RESULTS

The numbers were always normalised by the mean of the course i.e. $(clicks_in_A1_week > x)$ means $x\%$ of the mean value. For FP, the largest indicator of false prediction was increased activity in the week when the A1 was due. The pattern $(clicks_in_A1_week > 0.66)$ covers 908 students with confidence 86%. Adding activity one week before the deadline $(clicks_1week_before_A1 > 0.70)$ increases the confidence to 91%, covering 355 predictions. Lower activity in the last week supported with some activity in the week before adds 357 more predictions with confidence 77%. $(clicks_in_A1_week = [0.13;0.66]) \& (clicks_1week_before_A1 > 0.03)$. These two rules cover 1,063/1,610 = 66% of all FP with 84% confidence.

For FN, the only pattern reaching 80% confidence were students with activity that drops one week before the deadline of A1 and the week before, but only for courses with a high activity before the course start $(clicks_in_A1_week \leq 0.04) \& (\#studClicked_beforeStart > 0.56) \& (clicks_1week_before_A1 \leq 0.02)$. These covered 286 predictions with confidence 81%, i.e. 233 out of all 1,517 FN = 15%. Both for FP and FN, none of the important factors included demographics.

4 CONCLUSION AND FUTURE WORK

The identified patterns suggest that 66% highly at-risk students three weeks before the deadline can still exert activity and succeed in the assignment. On the other hand, FN are less interpretable and even the pattern covering 15% of FN needs closer investigation, particularly because it applies only for courses with many students who are active before the start of a course. Next steps for our work will include randomly selecting a fixed number of students from both FP and FN predictions that do not follow the pattern. These students will be interviewed, and thematic analysis will be conducted to provide more insight into phenomena not able to be captured by current data.

REFERENCES

- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114.
- Gardner, J., & Brooks, C. (2018). Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction*, 28(2), 127-203.
- Hlosta, M., Zdrahal, Z., & Zendulka, J. (2017). *Ouroboros: early identification of at-risk students without models based on legacy data*. Paper presented at the Proceedings of the Seventh International Learning Analytics & Knowledge Conference.
- Kitto, K., Shum, S. B., & Gibson, A. (2018). *Embracing imperfection in learning analytics*. Paper presented at the Proceedings of the 8th Intl. Conf. on Learning Analytics and Knowledge.
- Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., & Addison, K. L. (2015). *A machine learning framework to identify students at risk of adverse academic outcomes*. Paper presented at the Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining.
- Qiu, J., Tang, J., Liu, T. X., Gong, J., Zhang, C., Zhang, Q., & Xue, Y. (2016). *Modeling and predicting learning behavior in MOOCs*. Paper presented at the Proceedings of the ninth ACM international conference on web search and data mining.
- Springer, A., & Whittaker, S. (2018). "I had a solid theory before but it's falling apart": Polarizing Effects of Algorithmic Transparency. *arXiv preprint arXiv:1811.02163*.

Describing Teachers' Self-Regulation of Information Seeking: Preliminary Results from Physiological Arousals and Log Files

Lingyun Huang

Department of Educational and Counselling Psychology, McGill University
lingyun.huang@mail.mcgill.ca

Susanne P. Lajoie

Department of Educational and Counselling Psychology, McGill University
susanne.lajoie@mcgill.ca

ABSTRACT: The study examines teachers' self-regulation when they are seeking information in an open-ended learning environment. Multimodal data of log files and participants' electrodermal activities data during information seeking were collected and analyzed. The preliminary results from one participant's data demonstrated that teachers can execute self-regulated learning activities but might not maintain them throughout the problem-solving process. Teachers that regulate and orient their goals in an efficient way have relatively lower physiological arousal. Those with less self-regulation demonstrate aimless navigating behaviors and relatively higher physiological arousals. The findings could contribute to research into converging multimodal data to analyze SRL. However, more data are required for the interest of trustworthiness and generalizability.

Keywords: Multimodal learning analysis, Log files, Electrodermal activity, Physiological arousal, Teachers' regulation, Information Seeking

1. INTRODUCTION

Information seeking in open-ended learning environments (OELEs) is difficult since massive information that is non-linearly presented in OELEs might hinder teachers' seeking processes, leading to meaningless navigations. Success in information seeking requires teachers to orient navigations and monitor and adjust navigations. Success also requires teachers to manage time and affective states, especially when they have difficulties finding target information. Research indicates that self-regulated learning (SRL) is effective for information seeking since it accounts for metacognitive, behavioral, and affective regulating processes (Azevedo, Taub, & Mudrick, 2017). Self-regulated teachers are more aware of strategically monitoring their navigations (e.g., setting goals, etc.), allocating time appropriately, and controlling affective states while searching. Given the complexity of SRL processes, research embraces multimodal learning analytics (MMLA) techniques that converges multiple sources of data derived from online trace methods such as log files, physiological sensors to model or predict complex real-time interactions of metacognitive, behavioral, and affective regulatory processes (Azevedo & Gašević, 2019). In this work, we leverage log files and physiological arousal data to investigate teachers' SRL processes in online information seeking contexts. Previous work conceptually articulates how to use log files model SRL (e.g., Siadat, Gasevic, & Hatala, 2016) and empirically demonstrates the effectiveness of such a method (e.g., Taub & Azevedo, 2018). Physiological arousal in this study is measured by electrodermal activity (EDA) data that measures skin conductance responses to indicate the level of teachers'

physiological arousal while searching. Such data can predict when learners experience positive or negative feelings that are associated with learning performance (e.g., Harley, Jarrell, & Lajoie, 2019). Using MMLA, log files and EDA data are converged to indicate teachers' metacognitive, behavioral, and affective states when seeking information within an OELE. Since the full data analysis is still in progress, and this paper presents only preliminary results from a case study of one participant, Yun (anonymized name). With the case, we are interested in (1) if he oriented his information-seeking processes, monitored navigations, and evaluated the consequences of such navigations, and (2) if he had lower physiological arousal levels when he was in regulated processes.

2. METHODS, ANALYSIS, AND DISCUSSION

Participants were asked to design a lesson using nBrowser - an OELE wherein participants can analyze the tasks, search information online, and edit lesson plans (Poitras, Doleck, Huang, Li, & Lajoie, 2017). nBrowser provides hints to participants and records their interactions through log traces (e.g., reading a website). Participants wore a wearable EDA device, Q-Sensor2.0, on their left wrist throughout the entire process. The device captures the EDA at 4Hz. The first five minutes before the task was used to establish the baseline of skin conductance levels. Participants were then given 45 minutes to complete and submit their lesson plans.

Yun's log files extracted from nBrowser showed that he spent around 17 minutes seeking online information and that the time-stamped sequence was Navigation– Reading hints– Search– Return Home page – Search – Return Home page – Search – Navigation – Search – Navigation – Search. As nBrowser is designed based on SRL models (Winne & Hadwin, 1998), we can translate some events into SRL events, i.e., Reading hints -> help-seeking (HS.); Return Home page -> goals monitoring (GM.). We defined Navigation (Nav.) as actions of exploring the learning environments, which suggests a less relevant SRL trajectory. Consequently, Yun's regulated information-seeking process is described and presented in Figure 1 that includes single SRL processes and identified SRL patterns, as well as the corresponding time. Pattern 1, 2, and 3 represented good SRL processes, wherein Yun used different metacognitive skills to monitor and control his information-seeking behaviors, for example, help-seeking and goal monitoring. By contrast, Pattern 4 and 5 seemed less regulated, evidenced by more time on navigation and few metacognitive monitoring activities. The event descriptors in logs showed that Yun found target information after he completed the third SRL process. In the rest of the seeking, however, he visited different websites and searched for many topics but did not obtain results. In light of time allocation, Yun spent 6 minutes in Pattern 1, 2, and 3 while 11 minutes in Pattern 4 and 5, which could suggest he was more efficient in an early search stage and became less regulatory later. We added EDA data for further investigation. The EDA was standardized and interpreted between 0, a relatively low arousal response and 1, a relatively high response (J. M. Harley et al., 2019). The results indicated that Yun's average physiological arousals were relatively lower in the first three patterns but relatively higher in Pattern 4 and 5, suggesting that lower physiological arousal levels could be associative with self-regulation. Noticeably, Yun had extensive physiological responses in his late search (i.e., in Pattern 4 and 5). Relating the information from log files, we assume that Yun might be lost due to the lack of clear goals or search strategies. Another reason for higher arousal at the end of the session could be that he realized the task time limitations and consequently became more stressed.

SRL pattern (Time,s) EDA (M/SD)	Pattern 1 (150s) .636/.009			Pattern 2 (37s) .640/.009		Pattern 3 (202s) .662/.005		Pattern 4 (445s) .770/.775		Pattern 5 (23s) .838/.048	
SRL event (Time,s) EDA (M/SD)	Nav.(106s) .63/.016	HS.(145) .63/.009	Sear.(30s) .646/.019	GM.(28s) .647/.007	Sear.(9s) .633/.03	GM.(6s) .626/.001	Sear.(196s) .670/.038	Nav.(15s) .717/.010	Sear.(430s) .823/.083	Nav.(83s) .872/.027	Sear.(430s) .804/.017

Figure 1: The time and means and standard deviations of EDA for the SRL processes. Sear. = Search

3. CONCLUSIONS

In conclusion, we implemented the MMLA to investigate teachers' self-regulation in information-seeking activities. The preliminary results illustrate that log files and the EDA data can reflect teachers' SRL patterns in information seeking. When teachers are self-regulated, they could be efficient and effective in finding targets and experience lower physiological responses. In less regulated situations, teachers could feel lost and therefore undergo extensive arousal and become unproductive. The findings are in line with studies that use log files and EDA to analyze SRL in other disciplines (e.g., Taub & Azevedo, 2018). These findings provide an additional piece of evidence to articulate how teachers engage in self-regulation. However, since this is a preliminary analysis of one participant, the generalizability of the study findings is constrained. Moreover, the low sampling rate should be taken into consideration. Further analysis of the remaining data is needed to confirm or adjust the findings. The future direction could increase the sampling rate and include more data like facial expressions to validate teachers' affective states and changes in self-regulation.

REFERENCES

- Azevedo, R., & Gašević, D. (2019). Analyzing Multimodal Multichannel Data about Self-Regulated Learning with Advanced Learning Technologies: Issues and Challenges. *Computers in Human Behavior*, 96, 207–210. <https://doi.org/10.1016/j.chb.2019.03.025>
- Azevedo, R., Taub, M., & Mudrick, N. V. (2017). Understanding and Reasoning about Real-Time Cognitive, Affective, and Metacognitive Processes to Foster Self-Regulation with Advanced Learning Technologies. In D. H. Schunk & J. A. Greene, *Handbook of Self-Regulation of Learning and Performance* (2nd ed., pp. 254–270). <https://doi.org/10.4324/9781315697048.ch17>
- Harley, J. M., Jarrell, A., & Lajoie, S. P. (2019). Emotion regulation tendencies, achievement emotions, and physiological arousal in a medical diagnostic reasoning simulation. *Instructional Science*, 47(2), 151–180. <https://doi.org/10.1007/s11251-018-09480-z>
- Poitras, E., Doleck, T., Huang, L., Li, S., & Lajoie, S. P. (2017). Advancing teacher technology education using open-ended learning environments as research and training platforms. *Australasian Journal of Educational Technology*, 33(3). <https://doi.org/10.14742/ajet.3498>
- Siadat, M., Gasevic, D., & Hatala, M. (2016). Trace-based Micro-analytic Measurement of Self-Regulated Learning Processes. *Journal of Learning Analytics*, 3(1), 183–214. <https://doi.org/10.18608/jla.2016.31.11>
- Taub, M., & Azevedo, R. (2018). Using Sequence Mining to Analyze Metacognitive Monitoring and Scientific Inquiry based on Levels of Efficiency and Emotions during Game-Based Learning. *Journal of Educational Data Mining*, 10(3), 1–26. <https://doi.org/10.5281/zenodo.3554711>
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *The educational psychology series. Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Temporal analytics of log data derived from students' manipulating mathematical objects¹

Masataka Kaneko¹, Takahiro Nakahara², Takeo Noda¹

¹Toho University, ²Sangensha LLC.
masataka.kaneko@phar.toho-u.ac.jp

ABSTRACT: In this study, inquiry-based learning in college-level mathematics is investigated. Using iPad, students manipulated function graphs which were generated by dynamic geometry software. In our case study, we analyzed how students manipulated in an attempt to discover insights into the learning process and it was found that the change in sequential pattern of manipulation might illustrate the progress of understanding. However, positional information associated with each manipulation is also needed to correctly interpret students' thinking. Therefore, we implemented a Moodle-plugin by which students can manipulate mathematical object on the web and researchers can store and download the log data of manipulations including both temporal and positional information. Our pilot study shows that this Moodle-plugin can be a powerful tool that helps us make sense of learners' thinking process and get insights into educators' scaffolding strategies.

Keywords: Dynamic Geometry, Moodle, Log Data of Manipulations, Mathematics Education

1 INTRODUCTION

The nature of learning mathematics interactively and in a technology-rich environment has been investigated in many previous studies (Bookman & Malone 2003). The constructivist perspective predicts that inquiry-based learning with technology can facilitate students' mathematical thinking and make their understanding more stable. Among other technologies, dynamic geometry software has great potential since it enables students to manipulate mathematical objects like function graphs and intuitively grasp how mathematical theory works. However, it is not easy for teachers to understand, simply through observation, students' thinking. In fact, especially in cases when students try to solve ill-structured problems, their thinking process tends to become highly complex while some hidden efficacy can be expected (Kapur 2010). Thus, it will be helpful if teachers can make sense of students' manipulating process and find appropriate ways to scaffold students' understanding. In this paper, we present our case study which investigated students' inquiry of a fundamental concept in calculus by tracking behavior which reflects the qualitative change in their understanding as they manipulated function graphs. Based on the results, we show that the Moodle-plugin introduced in this paper will allow researchers to make fine-grained analysis and help teachers more easily interpret student behavior.

2 METHODS AND RESULTS OF CASE STUDY

¹ This work was supported by JSPS Kakenhi 18K02872

Approximation of function by Taylor polynomial is an important topic in college-level mathematics education. From our teaching experience, while it is not so difficult for students to memorize the formula and apply it to specific cases, they seem to fall into difficulty in appreciating its background mechanism and associated concepts like the order of the infinitesimal. To resolve this difficulty, we used a dynamic geometry system named CindyJS (<https://cindyjs.org>) to prepare HTML content including graphs of functions as in Figure 1(left). This content is presented on web browsers so that touch operation by students is enabled. Subjects were asked to manipulate sliders in the content and find the suitable coefficients to approximate the target function with cubic polynomial function near $x=0$. Unless the first order part is set to be the equation of a tangent line, any choice of higher order coefficients does not provide a suitable approximation. Observing this situation, students are expected to empirically understand the concept of the infinitesimal of higher order. The sequential pattern of students' manipulation should reflect the extent to which they become aware of this concept. Therefore, in our case study, we made a video recording of the iPad screens and then plotted their choices of sliders on one timeline per subject as shown in Figure 1(right).

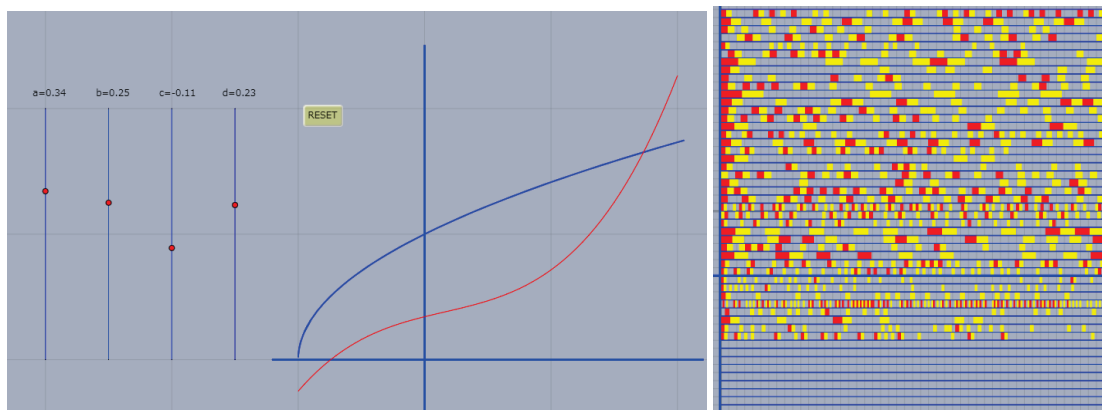


Figure 1: The content used (left) and the plot of subjects' manipulating process (right)

In the right figure, the horizontal axis pointing to the right represents the passage of time. Each subject's manipulation process is visualized in its entirety on the same length of interval, and each manipulation was given the same weight. Red and yellow correspond to the manipulation of zeroth order term and first order term respectively. The upper part (over the thick blue line) represents the log data of manipulations made by first-year university students of average ability ($N=33$) and the lower part represents those made by high school students of high academic ability ($N=8$). As seen in the figure, there seems to be a remarkable difference in the manipulation pattern at the later part of the manipulating process between the two groups of subjects.

3 DISCUSSION AND FUTURE DIRECTION

In our pilot study investigating the interrelation between students' manipulation of mathematical content and their discourse while they did this task in a group, it was found that, as students became aware of the target concept, they moved the sliders controlling the lower order coefficients less and less. The above result suggests that students' manipulating pattern depends on their academic ability and that therefore the appropriate way of scaffolding might vary accordingly. Thus, it can be effective to track the sequential pattern of students' activities; however, in the case of manipulating mathematical objects, this approach is insufficient because the positional information associated

with each manipulation is neglected. In fact, the same sequential pattern could generate an extremely wide range of graphical shapes. Since students should choose each manipulation based on their observations of the graphical shapes derived from their preceding trials, the whole manipulation process cannot be interpreted simply by sequential analysis. To study the positional information of students' manipulations, we implemented Moodle-plugin which enabled students to manipulate CindyJS content on the web and teachers to download the log data including positions of mathematical objects and corresponding time stamps. The log data are formatted into CSV file as in Figure 2(left). Using this output, we can compute various quantities associated with the log data and visualize the transition of those quantities. In fact, Figure 2(right) shows the transition of the radius of the “well approximated range” derived from the log data of one student's trial. This quantity seems to characterize students' manipulation strategy.

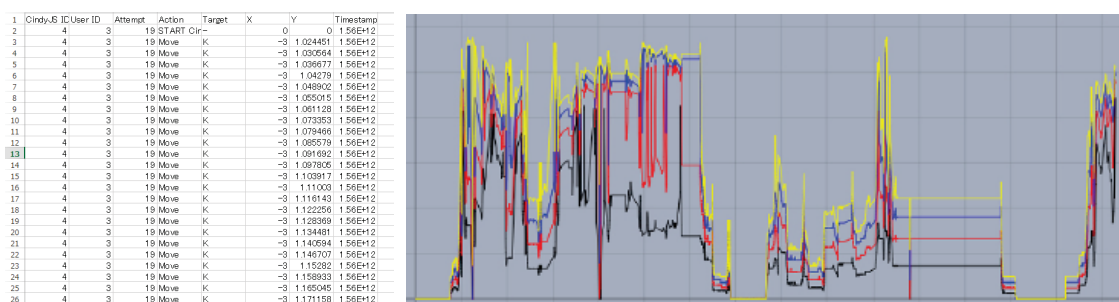


Figure 2: Log data (left) and the plot of the radius of “well approximated range” (right)

During the two time intervals in which the radius is equal to zero, the teacher gave some advice about the range of approximation. The figure on the right indicates that the teacher's advice influenced how the student approximated on the iPad and thus encouraging him to pursue other possibilities in later stages of his trials. As seen in this example, the visualization of the temporal transition of some characteristic quantity might demonstrate students' learning trajectory and give some insight into whether a teacher's specific intervention is appropriate or not.

The situation where fine-grained learning data and temporal analytics of the data are needed is quite similar to that of discourse analysis. As stated in a previous study (Mercer 2008), the nature of the shared knowledge is potentially quite complex since immediately shared experiences and corresponding conversational content provide the resources for future talk. In a sense, students' interaction with mathematical content is a kind of discourse. The newly developed work flow in this paper will uncover the streams of students' thinking implicitly embedded in seemingly chaotic pictures of learning data.

REFERENCES

- Bookman, J., Malone, D. (2003). The nature of learning in interactive technological environments: A proposal for a research agenda based on grounded theory. *Research in Collegiate Mathematics Education*, 5, 182-204.
- Kapur, M. (2010). Productive failure in mathematical problem solving. *Instructional Science*, 38, 523-550. <https://doi.org/10.1007/s11251-009-9093-x>
- Mercer, N. (2008). The seeds of time: Why classroom dialogue needs a temporal analysis, *The Journal of the Learning Sciences*, 17(1), 33-59. <https://doi.org/10.1080/10508400701793182>

Using network analysis to investigate the relation between knowledge organization and transfer

Marcus Kubsch

IPN – Leibniz Institute for Science and Mathematics Education, Kiel, Germany
kubsch@leibniz-ipn.de

ABSTRACT: Scientific literacy is an important part of education. Students that demonstrate scientific literacy can organize and coordinate their science ideas to interpret and explain a diverse range of phenomena. This often requires that students are able to transfer their knowledge to new contexts. While some general principles of knowledge transfer have been discovered, we still know little about what exactly enables experts in a field to transfer their knowledge while novices often fail to do so. However, a striking difference between experts and novices is how they organize their knowledge in a domain. We use network analysis to investigate how the organization of students' knowledge about energy influences their ability to transfer their knowledge to a new context. On this poster, we present first results, discuss their implication and sketch future directions towards automated analyses.

Keywords: Network analysis, knowledge transfer, knowledge organization

1 BACKGROUND

Regardless of where one looks, being able to use what one has learned in new contexts, i.e., being able to transfer one's knowledge, is emphasized as a major goal of learning, e.g., in the US *Framework for K-12 Science Education* (National Research Council, 2012). While the study of transfer has revealed some of the fundamental cognitive mechanisms that facilitate transfer such as analogical reasoning (Gick & Holyoak, 1983) or the importance of mastery goal orientation for succeeding in transfer tasks (Belenky & Nokes-Malach, 2012), the literature is also full of mixed results and it remains challenging to predict when and under what conditions students will be able to transfer their knowledge to new contexts – especially when it comes to discipline specific knowledge (J. Bransford & Schwartz, 1999). However, the study of expertise has identified that a key element of expertise is to apply one's discipline specific knowledge across a wide range of contexts. Further, the knowledge of experts in a domain is organized differently than the knowledge of novices – it is organized and strongly connected around key ideas in a domain (J. Bransford, 2000). Thus, how students organize their ideas in a domain should be related to their ability to transfer that knowledge. However, studies that investigated how students organize their knowledge, e.g., about energy (Lee & Liu, 2010), have only rarely connected their results to transfer. Similarly, while affective measures, especially goal orientation, have been found to be strong predictors of transfer, little is known about their relationship to the organization of students' knowledge networks. Therefore, we adopted a network analytical approach by Kubsch et al. (2019) to investigate the following research questions: 1) how is the organization of students' knowledge-networks about energy related to their ability to transfer their knowledge to a new context? 2) how is the organizations of students' knowledge networks related to their goal orientation?

2 METHODS

2.1 Design & sample

In this study, we draw interview data and results from a transfer task that were part of a larger study that investigated the learning of energy in physics in middle school. The interview data is needed to construct knowledge networks for each student and relating those to the results from the transfer task allows us to answer research question 1). While the subsample that was interviewed and took the transfer task is small ($N=20$) it is representative of the sample as a whole ($N=394$), based on a test of students' energy understanding. Further, we draw on a goal orientation measure adapted from Vedder-Weiss & Fortus (2011), to address research question 2).

2.2 Instruments

The interviews followed an interview-about-instances approach (Osborne & Gilbert, 1980) in which students were presented with different phenomena (5) and asked to explain them. Interviewers did not prompt students to use science ideas and follow-up questions were limited to clarification questions using the language of the student to avoid directing or leading students in any way.

The transfer task took about 45 minutes and was centered on the topic of reusable hand warmers. Students had to answer two open ended question that required students to engage in the scientific practices of modelling and argumentation and thus provided rich evidence of their ability to transfer their knowledge about energy from physics to chemistry.

The goal orientation measure consisted of 12 items with a 5 point likert-scale and measured to what extent students were mastery oriented.

2.3 Analyses

Following the approach by Kubsch et al. (2019), we coded the normative science ideas that students used in the explanations of the phenomena as the basis for constructing students' knowledge networks. We calculated network coherence to quantify the overall integratedness of students' knowledge networks and calculated the measure of degree for different energy ideas to assess their relative importance. The transfer task was scored by experienced scorers and we calculated sum scores. In both analyses, satisfactory inter-rater agreement was found. Lastly, the goal orientation measure showed satisfactory reliability.

3 RESULTS

For RQ1 we found a statistically significant correlation between students' network coherence and their transfer task score. ($.54, p < .05$). Further, we found that the degree of energy transfer and energy transformation ideas was relatively strongly correlated with students' scores ($.43, p = .05$ / $.37, p < .10$) while the network degree of the energy forms idea was only weakly correlated ($.23, p > .10$). In sum, this indicates that having overall well-connected, that is integrated, knowledge-networks around energy is most important for successful transfer and that within these networks the ideas of transfer and transformation are more important than the idea of forms. For RQ2 we found a large

statistically significant correlation between students' mastery goal orientation and the overall integratedness of their knowledge networks (.48, $p < .05$).

4 DISCUSSION & OUTLOOK

Our results indicate that having strongly connected knowledge networks around energy transfer and transformation ideas is strongly related with one's ability to transfer energy ideas to a new context. Further, a desire for understanding, i.e., mastery goal orientation, goes hand in hand with having well-connected knowledge networks. This is not only in line with previous findings (Belenky & Nokes-Malach, 2012) that related mastery learning to high transfer scores but goes beyond them as it hints at the mechanism: having a stronger mastery orientation, may lead to cognitive processes that results in better connected and organized knowledge networks. Further, these results provide empirical evidence for the widely held assumption (Wagner, 2006) that the details of how domain specific knowledge is organized play an important role in the ability to transfer. While our current study is limited by the small sample, which is a natural consequence of the interview approach, learning analytics techniques like NLP and automated analyses can help to scale the approach in future studies. This also opens the possibility for a unique kind of feedback that one could provide to students. In addition to telling them what they can already do and what they still have to learn, one could point connections between ideas that they may be missing and direct them to activities that can help to build these connections. Further, the importance of mastery goal orientation, also suggests interventions that help students become more mastery oriented. Currently, we are working to implementing these ideas in a digital learning environment that is developed in a new project.

REFERENCES

- Belenky, D. M., & Nokes-Malach, T. J. (2012). Motivation and Transfer: The Role of Mastery-Approach Goals in Preparation for Future Learning. *Journal of the Learning Sciences*.
- Bransford, J. (2000). *How people learn: Brain, mind, experience, and school* (Expanded ed; National Research Council (U.S.), Ed.). Washington, D.C: National Academy Press.
- Bransford, J. D., & Schwartz, D. L. (1999). Rethinking Transfer: A Simple Proposal With Multiple Implications. *Review of Research in Education*, 24(1), 61–100.
- Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*.
- Kubsch, M., Nordine, J., Neumann, K., Fortus, D., & Krajcik, J. (2019). Probing the Relation between Students' Integrated Knowledge and Knowledge-in-Use about Energy using Network Analysis. *Eurasia Journal of Mathematics, Science and Technology Education*, 15(8).
- Lee, H.-S., & Liu, O. L. (2010). Assessing learning progression of energy concepts across middle school grades. *Science Education*, 94(4), 665–688.
- National Research Council. (2012). *A framework for K-12 science education*. Osborne, R. J., & Gilbert, J. K. (1980). A technique for exploring students' views of the world. *Physics Education*, 15(6), 376.
- Vedder-Weiss, D., & Fortus, D. (2011). Adolescents' Declining Motivation to Learn Science: Inevitable or Not? *Journal of Research in Science Teaching*, 48(2), 199–216.
- Wagner, J. F. (2006). Transfer in Pieces. *Cognition and Instruction*, 24(1), 1–71.

Accurate and Interpretable Sensor-free Affect Detectors via Monotonic Neural Networks

Andrew S. Lan¹, Anthony Botelho², Shamy Karumbaiah³, Ryan S. Baker³, Neil Heffernan²

¹University of Massachusetts Amherst, ²Worcester Polytechnic Institute, ³University of Pennsylvania

ABSTRACT: Sensor-free affect detectors can detect student affect using their activities within intelligent tutoring systems or other online learning environments rather than using sensors. This technology has made affect detection more scalable and less invasive. However, existing detectors are either interpretable but less accurate (e.g., classical algorithms such as logistic regression) or more accurate but uninterpretable (e.g., neural networks). We investigate the use of a new type of neural networks that are *monotonic* after the first layer for affect detection that can strike a balance between accuracy and interpretability. Results on a real-world student affect dataset show that monotonic neural networks achieve comparable detection accuracy to their non-monotonic counterparts while offering some level of interpretability.

Keywords: Affect detection, interpretability, neural networks

1 INTRODUCTION

Affect detectors that can detect and monitor student affective states have become an important aspect of learning analytics research. Together with methods that can trace students' knowledge levels over time, they can support timely and personalized interventions to improve student learning outcomes. Existing student affect detection methods can be classified into two classes. One class employs physical and physiological sensors to measure students as they learn, which is accurate but invasive and not scalable, the other "sensor-free" class uses machine learning-based classifiers to detect a student's affective state from their recorded activity in the ITS, which is non-invasive, scalable, but is in some cases less accurate [Bosch et al., 2015; Henderson et al., 2019]. The trade-off a sensor-free affect detector achieves in terms of accuracy and interpretability is closely related to the type of classification algorithm it uses. Detectors based on classic algorithms such as logistic regression, i.e., [Pardos et al., 2014] can be more interpretable but less accurate, while neural network-based detectors can be more accurate but not interpretable [Botelho et al., 2017]. Therefore, there is a need to develop new classifiers that can find better trade-offs between accuracy and interpretability; we propose to use monotonic neural networks as a potential solution.

2 MONOTONIC (FULLY-CONNECTED) NEURAL NETWORKS

For sensor-free affect detection, we are given a student activity feature vector $\mathbf{x} \in \mathbb{R}^K$, where K denotes the number of features used to summarize student activities within a learning system during an affect observation, and our goal is to detect whether or not a student is in a certain affective state y , which is (typically) binary-valued. Affect detectors are typically classifiers such as logistic regression

$$p(y = 1) = \sigma(\mathbf{w}^T \mathbf{x}) = 1/(1 + e^{-\mathbf{w}^T \mathbf{x}}),$$

where $\mathbf{w} \in \mathbb{R}^K$ denotes the regression coefficient (bias is omitted for simplicity of exposition). The values of regression coefficients offer us excellent interpretability since they explicitly control the probability of the student being in this affective state via a linear relationship. Other classic algorithms such as decision trees offer reasonably high interpretability as well, e.g. [Paquette et al., 2014]. Recent research has suggested that neural networks can often achieve significantly better predictive accuracy

than logistic regression for binary classification problems [Goodfellow et al., 2016]. In this paper, we use fully connected neural networks to improve the accuracy of affect detection. However, these detectors are often uninterpretable due to the presence of multiple layers and nonlinearities. In order to add interpretability to these neural networks, we propose to investigate the family of “monotonic” neural networks by i) selecting monotonic activation functions and ii) restricting weights beyond the first layer to be nonnegative. We note that common nonlinearities are monotonic, such as hyperbolic tangent (*tanh*) and rectified linear units (*ReLU*) [Goodfellow et al., 2016]. Using a two-layer neural network as an example, for hidden unit i in the first layer, we have

$$p(y = 1) = \sigma(W_{2,i} z_i + \text{const}) = \sigma(W_{2,i} \Phi(\mathbf{w}_i^T \mathbf{x}) + \text{const}),$$

where Φ denotes the nonlinearity in the first layer, z_i denotes the value of this hidden unit, and $W_{2,i}$ denotes the weight in the second layer connecting this hidden unit to the output. It is easy to show that when Φ is monotonic and $W_{2,i}$ is nonnegative, the probability of a student being in this affective state is also monotonic with respect to $\mathbf{w}_i^T \mathbf{x}$, a property shared with logistic regression. This observation can be generalized to multi-layer neural networks and enable us to interpret neural network-based affect detectors using the coefficient \mathbf{w}_i for each hidden unit in the first layer, if weights in subsequent layers are nonnegative. Despite the presence of nonlinearities at each layer preventing us from comparing the relative importance of features using their coefficients, we can still conclude that whether a feature is positively or negatively correlated with an affective state.

3 EXPERIMENTS

We conduct a series of experiments using monotonic networks as affect detectors on the ASSISTments student affect dataset¹, which was collected in real classrooms as students work within the ASSISTments system by observers following the Baker Rodrigo Ocumpaugh monitoring protocol (BROMP) [Ocumpaugh et al., 2015]. The dataset contains 3109 observations. Each observation contains i) a student's affective state label during a 20-second observation interval and ii) a set of 88 features that summarizes their activities within ASSISTments during this time interval. A total of 4 affective states were coded in this data set: bored, confused, engaged concentration, and frustrated. In this paper, we only analyze the detection of engaged concentration, since it is the most common.

We separate the entire dataset into a training set with 70% of the observations, a validation set with 10% of the observations, and a test set with 20% of the observations. We test four different detectors using four different classifiers: logistic regression (LR), random forest (RF), fully-connected neural network (FNN), and its monotonic version (M-FNN). For each detector, we use the validation set to select the best parameter setting and report detection performance on the test set. For the neural network-based detectors, we sweep over algorithm parameters as learning rate $\in \{1e-5, 1e-4, 1e-3\}$, number of layers $\in \{2, 3\}$, number of units in each layer $\in \{5, 10, 20\}$, nonlinearity $\in \{\text{tanh}, \text{ReLU}\}$, and different random initializations of the network weights and biases. For the LR and RF detectors, we sweep over the learning rate and number of decision tree parameters, respectively, using a similar approach.

Table 1 shows the performance of each affect detector on the test set, with means and standard deviations calculated over 10 random partitions of the dataset. We see that neural network-based detectors significantly outperform LR- and RF-based detectors, and the monotonic version of the FNN-based detector achieves similar performance to that of its unrestricted version. Table 2 shows the top features and corresponding (regression) coefficients for most predictive features in the LR and M-FNN detectors (we selected one hidden unit in the hidden unit for the latter). We see that the top features (not coefficient values) match up reasonably closely across both cases.

¹ This dataset is taken from <http://tiny.cc/affectdata>.

Table 1: Engagement detection accuracy on the ASSISTments dataset for all detectors compared.

	AUC
LR	0.746 \pm 0.036
RF	0.763 \pm 0.029
FNN	0.782 \pm 0.030
M-FNN	0.780 \pm 0.032

Table 2: Most predictive features for engagement in the LR and M-FNN (1 unit) detectors.

LR		M-FNN	
Feature	Coefficient	Feature	Coefficient
max_frWorkingInSchool	-0.101	max_frWorkingInSchool	-0.471
min_correct	0.096	avg_stlHintUsed	-0.420
avg_hintTotal	-0.066	sum_hintCount	-0.379
sum_timeTaken	-0.054	sum_timeTaken	-0.369
avg_stlHintUsed	-0.032	avg_frPast8WrongCount	-0.359

4 LIMITATIONS AND FUTURE WORK

Though this approach increases interpretability, we have found that we can only interpret the directionality of each unit in the first hidden layer of the neural network separately. Moreover, our monotonic restrictions do not apply to recurrent neural networks e.g., [Botelho et al., 2017] since these restrictions would enforce monotonicity on affect over time as well as activity features. Finally, we have not yet established if similar patterns would hold for other, less frequent affective states.

REFERENCES

- Bosch, N., Chen, H., Baker, R. S., Shute, V., D'Mello, S. (2015). Accuracy vs. Availability Heuristic in Multimodal Affect Detection in the Wild. In *Proc. International Conference on Multimodal Interaction*, 267-274.
- Botelho, A. F., Baker, R. S., & Heffernan, N. T. (2017). Improving sensor-free affect detection using deep learning. In *Proc. International Conference on Artificial Intelligence in Education*, 40-51.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT press.
- Henderson, N., Rowe, J., Mott, B., Brawner, K., Baker, R. S., Lester, J. (2019). 4D Affect Detection: Improving Frustration Detection in Game-based Learning with Posture-based Temporal Data Fusion. In *Proc. International Conference on Artificial Intelligence in Education*, 144-156.
- Ocuppaugh, J., Baker, R. S., & Rodrigo, T. (2015). Baker Rodrigo Ocuppaugh monitoring protocol (BROMP) 2.0 technical and training manual. *New York, NY and Manila, Philippines: Teachers College, Columbia University and Ateneo Laboratory for the Learning Sciences*, 60.
- Paquette, L., Baker, R. S., Sao Pedro, M. A., Gobert, J. D., Rossi, L., Nakama, A., & Kauffman-Rogoff, Z. (2014). Sensor-free affect detection for a simulation-based science inquiry learning environment. In *Proc. International Conference on Intelligent Tutoring Systems*, 1-10.
- Pardos, Z. A., Baker, R. S., San Pedro, M. O., Gowda, S. M., & Gowda, S. M. (2014). Affective states and state tests: Investigating how affect and engagement during the school year predict end-of-year learning outcomes. *Journal of Learning Analytics*, 1(1), 107-128.

Online Learning Diagnosis and Feedback System Using Academic Emotion Data

Jihyang Lee

Ewha Womans University
jihyang@ewha.ac.kr

Hyo-Jeong So

Ewha Womans University
hyojeongso@ewha.ac.kr

ABSTRACT: In this paper, we firstly present the main ideas of designing an online learning diagnosis and feedback system that aims to provide affective feedback based on the analysis of academic emotion data. Then, we present the results of the experiment with 10 adult learners watching a MOOC video clip. The experiment study reveals a set of academic emotions that includes 7 types of positive and 11 types of negative academic emotions, which were detected from the analysis of various facial expressions. As future research steps, we propose the need to analyze co-occurring multiple emotional statuses and to build robust modeling coupled with bio-signal data.

Keywords: Emotional Computing, Academic Emotions, Online Learning System

1 INTRODUCTION

Support for learning in the emotional domain can create a friendly and enjoyable learning experience that can increase learner motivation and reduce exogenous cognitive loads. The growing awareness of the importance of affective learning has raised the necessity of studying academic emotions. Affective computing platforms (e.g., Affectiva) identify online learners' emotions by comprehensively analyzing facial expressions according to learners' emotional change, gestures, and various biometric data. Such analytical data can be the basis for providing affective feedback to support learners' motivation and volition, which can have a positive effect on learners' commitment to continue online learning (Dowell & Graesser, 2014). A high dropout rate is another critical problem in online learning. While various reasons exist, one critical reason is the difficulty of regulating the learning process in a massive online environment. With this backdrop, this paper is based on the research project that aims to design an online learning diagnosis and feedback system that utilizes academic emotion data to provide affective feedback that helps learners self-regulate learning processes. In this work-in-progress paper, we firstly present the main ideas of designing the proposed system and then discuss the analysis of learners' academic emotions in an online learning situation.

2 ONLINE LEARNING DIAGNOSIS AND FEEDBACK SYSTEM

Figure 1 shows the schematic diagram of the online learning diagnosis and feedback system proposed in this study. A learner interacts with the 'guiding avatar' that is an intelligent virtual agent to induce a specific emotion according to scenarios. In addition, emotion-related signals in response

to external stimuli are transmitted through biometric devices such as an electrocardiogram (ECG), electromyogram (EMG), body temperature (ST), and skin electrical conductivity (SCR) detected through a wearable bio-band device. Measured by the sensor, the learner's emotional state is recognized through noise filtering and pattern matching. Concurrently, the emotional state of the learner is statistically compared with the accumulated data of other learners who have learned the same content to determine the learners' current emotional state. Compared to other methods that rely on self-reported data to measure academic emotions, the proposed system provides both qualitative and quantitative data to better assist instructors and other users in making objective and data-driven diagnoses. Adopting mixed data can compensate for the shortcomings of individual data by giving better understanding and objectivity in context.

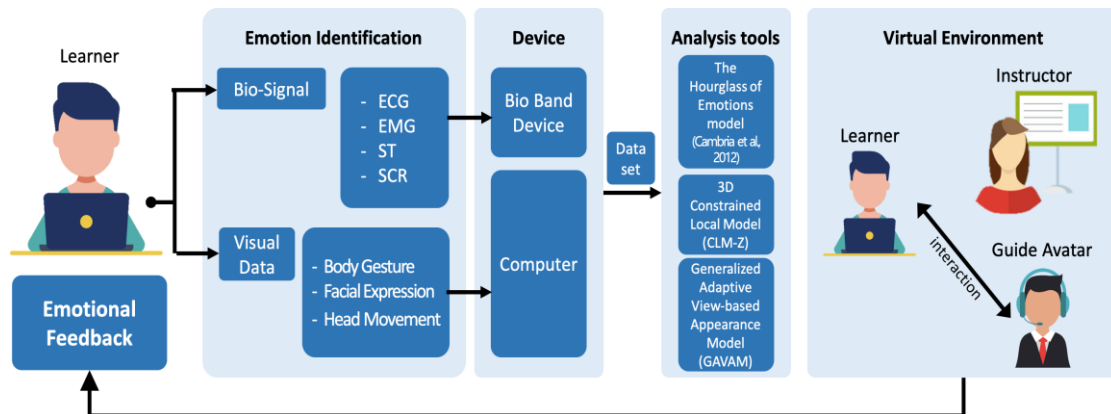


Figure 1: Online Learning Diagnosis and Feedback System

3 CLASSIFICATION OF ACADEMIC EMOTIONS

While previous studies have proposed the classification of academic emotions, little attempts have been made to understand academic emotions in video-based online learning situations. Hence, to build the proposed system, it is necessary to model academic emotions expressed in the online learning process, which is the experimental study that we conducted as an initial step of modeling.

3.1 Data Collection and Analysis

The participants were 10 adult learners in their 20s purposely recruited for the study. They were asked to watch a video lecture on physics from K-MOOCs (Korea-Massive Open Online Courses). The experiment was conducted for 90 minutes, and the whole learning process was video-recorded and some clips were used for a recall simulation interview conducted after completing video-based online learning. The analysis went through two phases. We firstly performed manual open coding based on a bottom-up approach and then utilized Nvivo for a systematic analysis of academic emotions from the captured data.

3.2 Results

Table 1 shows the results of classifying the types of academic emotions according to facial expressions generated during the online learning process. Due to the space constraint, we present only partial data related to mouth expressions. First, we identified 13 types of facial expressions (close lips, murmur, big breath, laugh, etc.) associated with 7 types of positive academic emotions: immersed, comprehending, interesting, curious, empathic, accepting, intentional and relieved. Second, 13 types of facial expressions (e.g., stick out lips, lip bite, big breath, yawning, rolling eyes, etc.) aroused from 11 types of negative academic emotions (e.g., distraction, displeasure, tension,

boredom, indifference, embarrassment, etc.). The emotions were classified based on the framework by Pekrun (2011).

Table 1: Classification of online academic emotion types by facial expressions

Type	Facial expression		Academic emotion	Frequency	%
Positive	Mouth	Close lips	Immersed / Comprehending/ Interesting / Intentional	52	22.03
		Murmur	Immersed / Empathic / Intentional	13	6.77
		Big breath (short)	Immersed / Empathic	4	1.69
		Laugh	Empathic / Interesting / Accepting	22	9.32
Negative	Mouth	Stick out lips	Offensive (opinion) / Objection / Curiosity	11	5.28
		Lip bites	Discomfort (content) / Tension	13	6.25
		Big breath (long)	Boredom / Immersive Release / Indifference	10	4.80
		laugh	Embarrassment (content)	4	1.92
		yawn	Immersive Release	9	4.32

4 CONCLUSION AND FUTURE STEPS

During the analysis process of the above experiment, we found that as multiple emotions occur, it is necessary to detect and analyze the flow of emotional changes and co-occurring multiple emotional statuses. Human emotions are complicated and vulnerable to situational factors (Harley, Bouchet, & Azevedo, 2012). Hence, there is a deep difference even in the same category of emotions. Recognizing such complex nature of human emotions, Cambria, Livingstone, and Hussain (2012) proposed the “Hourglass of emotions” model that includes four independent emotions and 32 concomitant dimensions of emotions. Hence, our next step of academic emotion modeling is to use this model to further analyze the full range of academic emotions that occur as single or multiple emotional status. Further, biometric data using a bio-band will be collected to triangulate data captured from the academic emotion analysis and various bio-signal data to build robust emotion modeling that will be shared at the conference.

REFERENCES

- Cambria, E., Livingstone, A., & Hussain, A. (2012). The hourglass of emotions. In *Cognitive behavioural systems* (pp. 144-157). Springer, Berlin, Heidelberg.
- Dowell, N. M. M., & Graesser, A. C. (2014). Modeling learners’ cognitive, affective, and social processes through language and discourse. *Journal of Learning Analytics*, 1(3), 183-186.
- Harley, J. M., Bouchet, F., & Azevedo, R. (2012, June). Measuring learners’ co-occurring emotional responses during their interaction with a pedagogical agent in MetaTutor. In *International Conference on Intelligent Tutoring Systems* (pp. 40-45). Springer, Berlin, Heidelberg
- Pekrun, R. (2011). Emotions as drivers of learning and cognitive development. In *New perspectives on affect and learning technologies* (pp. 23-39). Springer, New York, NY.

Triangulating Multimodal Data to Measure Self-Regulated Learning

Author(s): Lim¹, K.P., van der Graaf², J., Fan³, Y., Engelmann¹, K., Bannert¹, M., Molenaar², I., Gasevic³, D., Moore³, J.

Technical University of Munich¹

Radboud University²

University of Edinburgh³

lyn.lim@tum.de

ABSTRACT: The poster presents the first study of a project which focus on improving measurement and real-time support of self-regulated learning (SRL) using learning analytics. Education is increasingly focused on students' ability to regulate their own learning within technology-enhanced learning environments. Current SRL interventions do not sufficiently adapt to the individual learning process, thus, learning analytics offer an approach to better understand SRL-processes. As current approaches lack validity or require extensive analysis after the learning process, we aim to investigate how to advance support given to students by 1) improving unobtrusive data collection and machine learning techniques to gain better measurement and understanding of SRL-processes and 2) using these new insights to facilitate students' SRL by providing personalized scaffolds. We will reach this goal with a series of exploratory, lab, and field studies. The setup presented here consisted of a learning environment presented on a computer with a screen-based eye-tracker. Other data sources are log files, screen recording, and audio of students' think aloud. The analysis will focus on aligning the different data sources and detecting sequences that are indicative of micro-level SRL processes as a stepping stone for improving real-time scaffolds.

Keywords: self-regulated learning; personalized scaffolds; learning analytics; machine learning; adaptive systems

1 TRACE DATA TO DETECT SELF-REGULATED LEARNING PROCESSES

This project (funded by ORA; BA20144/10-1, NWO 464.18.104, ES/S015701/1) aims to improve measurement of self-regulated learning (SRL) by using multimodal learning analytics. SRL occurs when learners monitor and regulate content they access and operations they apply to operate on content as they pursue goals to augment and edit prior knowledge (Winne, 2019). Previous studies have shown that SRL is related to better learning outcomes and interventions can improve SRL and learning outcomes (e.g. Bannert & Reimann, 2012). However, the need for improved SRL measures has increased to capture processes while they occur (Schunk & Greene, 2018) as there is still no agreement to the appropriate learning actions to measure, diagnose, understand, and support students' SRL (Papamitsiou & Economides, 2014). A solution is to assess SRL at a more fine-grained level by measuring micro-level SRL processes. Unobtrusive measures of SRL can be captured through trace data in digital learning environments. Such traces are less biased than self-reports due to their temporal proximity (Gasevic, Jovanovic, Pardo, & Dawson, 2017), but traces do not reflect SRL processes on their own (Molenaar & Järvelä, 2014). Think aloud data has shown to be more insightful in determining SRL activities and predicting students' learning achievements than self-reports (Bannert, 2007). The integration of trace data with think aloud data provides opportunities to better measure micro-level SRL processes.

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

2 MEASURING MICRO-LEVEL SRL PROCESSES

This project will investigate and improve log data in two exploratory studies and develop and test personalized scaffolds based on individual learning processes in two laboratory and one subsequent field study. All studies are set in a scenario in which students are tasked to learn about artificial intelligence, differentiation, and scaffolding, and to write an essay, see Fig. 1 (left) for a schematic overview of the exploratory studies. Before and after this task, students' knowledge about the topics is assessed. Preliminary results of the first study show that there is a significant learning gain. How this learning gain is related to micro-level SRL processes will be investigated in the exploratory studies. Three types of data will be gathered: think aloud (audio), log data (mouse and keyboard), and eye-tracking (gaze).

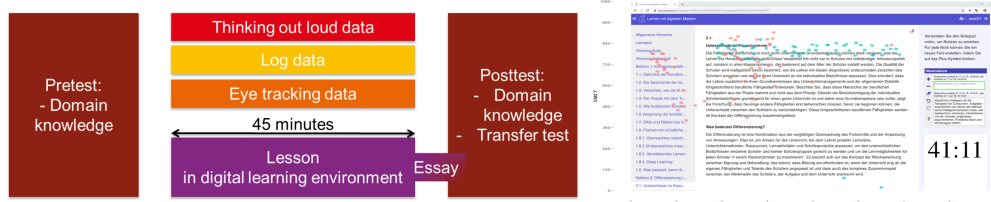


Figure 1: Study design (left) and the digital learning environment with eye fixations (right).

Think aloud can be coded to detect cognitive and metacognitive processes (micro-level SRL processes). But it is unclear if and how this process presents itself in log data or eye-tracking, see Fig. 1 (right). Think aloud data has been shown to be a good indicator of SRL processes and past research has paved a clear path for its analysis. The relation between SRL processes and the other data sources, however, is less evident. To tackle this, think aloud data will be used as an indicator of SRL and other data sources as proxies of SRL processes. In particular, sequential patterns are expected to be indicative of SRL processes. The challenge is to triangulate individual data streams and find proxies of micro-level SLR processes in log data.

2.1 Experimental setup and data analysis approach

The following setup were used: Screen-based eye-trackers (Tobii Pro Spectrum/TX300), webcams with microphones, keyboards, and mice. Data were collected synchronously on a computer using iMotions/Tobii Studio while the learning environment was presented. Multiple data sources were configured and combined into a log file. *Audio* was recorded to measure think aloud data—used for coding SRL processes. Previously developed and validated coding schemes for SRL were used to score the think aloud data (Bannert, 2007). *Log data* (mouse and *keyboard* data) indicate how the participant interacted with the learning environment. *Eye tracking data* was sampled at 300 Hz and consisted of fixations, saccades, gaze points, pupil size etc.

The data analysis was conducted in two steps: (a) We developed a trace parser which processed the trace data by first labeling all raw trace data to form an action library. A collection of actions formed meaningful learning patterns (i.e. a pattern library). These patterns formed the overall categories (e.g. Planning) in order to map to the think aloud data. (b) Next steps include the combination of multiple data streams.

2.2 Preliminary results

We segmented and coded think aloud protocols. To demonstrate the alignment of data sources, periods of orientation (metacognition) and reading (cognition) were contrasted, see Table 1. The difference in codes for the think aloud and log files showed that orientation was related to interactions with the menu (i.e. navigation), while reading did not leave any traces in the log data. For eye-tracking, the distance between subsequent fixations was larger for orientation ($M = 105.57$, $SD = 65.84$) compared to reading ($M = 79.12$, $SD = 41.16$), $t(110) = 2.54$, $p = .013$, $d = 0.48$.

Table 1: Triangulating multimodal data of SRL events to think aloud

Data source	Observations and possible processes from multiple data streams	
Think aloud	Orientation (<i>Task analysis</i>)	Reading (<i>Reading content</i>)
Log files	Navigation (<i>Task overview</i>)	None (<i>Accessing content</i>)
Eye tracking	Large distance (<i>Overview</i>)	Small distance (<i>Details</i>)

3 DISCUSSION AND NEXT STEPS

Using think aloud data to shed light on the measurement of SRL processes with multimodal data facilitates our understanding on how, when, and for whom to provide real-time support during SRL. In the example above, navigation actions and fixation patterns in eye-tracking data appeared to be an indicator for orientation. This shows how learning analytics can be applied in SRL research to better understand the learning process and ultimately, support SRL. The results are a stepping stone that demonstrate how meaning can be uncovered in multimodal data through the use of think aloud data as the ground truth. The project continues by developing an advanced algorithm to analyze learning processes and test its application in authentic learning settings.

REFERENCES

- Bannert, M. (2007). *Metakognition beim Lernen mit Hypermedia*. Münster, Germany: Waxmann
- Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through prompts. *Instructional Science*, 40, 193–211.
- Gasevic, D., Jovanovic, J., Pardo, A., & Dawson, S. (2017). Detecting Learning Strategies with Analytics: Links with Self-reported Measures and Academic Performance. *Journal of Learning Analytics*, 4(2), 113–128–113–128.
- Molenaar, I., & Järvelä, S. (2014). Sequential and temporal characteristics of self and socially regulated learning. *Metacognition and Learning*, 9(2), 75–85.
- Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Journal of Educational Technology & Society*, 17, 49–64.
- Schunk D.H., & Greene J.A. (2018). *Handbook of Self-Regulation of Learning and Performance (2nd ed.)*. London, UK: Routledge.
- Winne, P.H. (2019). Paradigmatic Dimensions of Instrumentation and Analytic Methods in Research on Self-Regulated Learning. *Computers in Human Behavior*, 96, 285–289.

How do the Game Level Plateaus Inform the Learning Design?

Anna Lizarov

Teachers College Columbia University
al3868@tc.columbia.edu

Charles Lang

Teachers College Columbia University
charles.lang@tc.columbia.edu

Kara Carpenter

Teachley, LLC
kara@teachley.com

ABSTRACT: Learning analytics can serve as a powerful tool for the evaluation of learning design and allow for data-driven and evidence-based modifications of the design of learning activities (Jayashanka, Hewagamage, & Hettiarachchi, 2018; Nguyen, Hupitych, & Rienties, 2018). However, there is a lack of research examining the intersection of learning design and educational digital math games. In this paper, we sought to examine the role of digital math games in the learning design of grade-school math curriculum. We investigated how the learning design incorporating the digital fractions game in the third-grade math curriculum in 2018 informed the learning design of such in 2019 by analyzing the game flow and level plateaus based on the number of replays per game level. We conclude that revisions made in the 2019 math curriculum based on 2018 gameplay data did not lead to an increase in students' learning gains.

Keywords: Learning Design, digital fractions math game, game level plateaus, Sankey diagram.

1 INTRODUCTION

Learning design is defined as a “methodology for enabling teachers/designers to make more informed decisions in how they go about designing learning activities and interventions, which is pedagogically informed and makes effective use of appropriate resources and technologies” (Conole, 2012, p. 7). In a technologically-driven age, an increasing number of classrooms are implementing educational games in their curriculum to engage students and deliver learning content. Recently, there has been a growing interest in applying Learning Analytics methods in analyzing game trace data, known as the Game Learning Analytics (Alonso-Fernandez et al., 2017). The game’s replayability feature engages students in the loop of making a decision based on the feedback, take action, and receive feedback based on this action. Gunter, Kenny, and Vick (2008) argued that educational games’ replayability is a critical feature for making knowledge and skills automatic, thereby, permitting higher-order thinking. Peddycord-Liu et al. (2017) found that certain elective replay behaviors are correlated with learning while other patterns suggest which educational content the students might be struggling with. Hence, elective replay behavior can be useful in providing intervention in a timely manner (Peddycord-Liu et al., 2017). Unfortunately, there are no studies examining the effect of the number of level replays on the learning outcome. The purpose of this study was to examine how incorporating

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

a digital mathematics game into a third-grade math curriculum in 2018 informed the learning design of that curriculum in 2019 by investigating the game level plateaus. Specifically, we examined whether revisions made to 2019 learning design resolved the issues associated with the game level plateaus found in 2018.

2 METHODS

The data has been retrieved from a digital fractions game, which consists of 12 levels. The data has been collected in a third-grade classroom during the time periods of May 2018 – June 2018 and February 2019 – March 2019. The 2018 school year sample consisted of 41 third-grade students, while there were 25 students participating in the research in the 2019 school year. During the 2018 school year, the gameplay took place in the months of May and June for six weeks in a classroom after explicit instruction on fractions was conducted. In 2019, however, the gameplay was aligned with the “fractions” topic in the curriculum, which was in February and March. The students had access to the game at home, after school, and after the curriculum intervention. Data was collected on levels completed, the number of replays per level, and the number of problems solved per level for each student. Replay patterns were identified using Sankey diagrams and effects tested for with a two-sample Wilcoxon rank-sum test and Vargha and Delaney’s *A* was used for effect size measure.

3 RESULTS & DISCUSSION

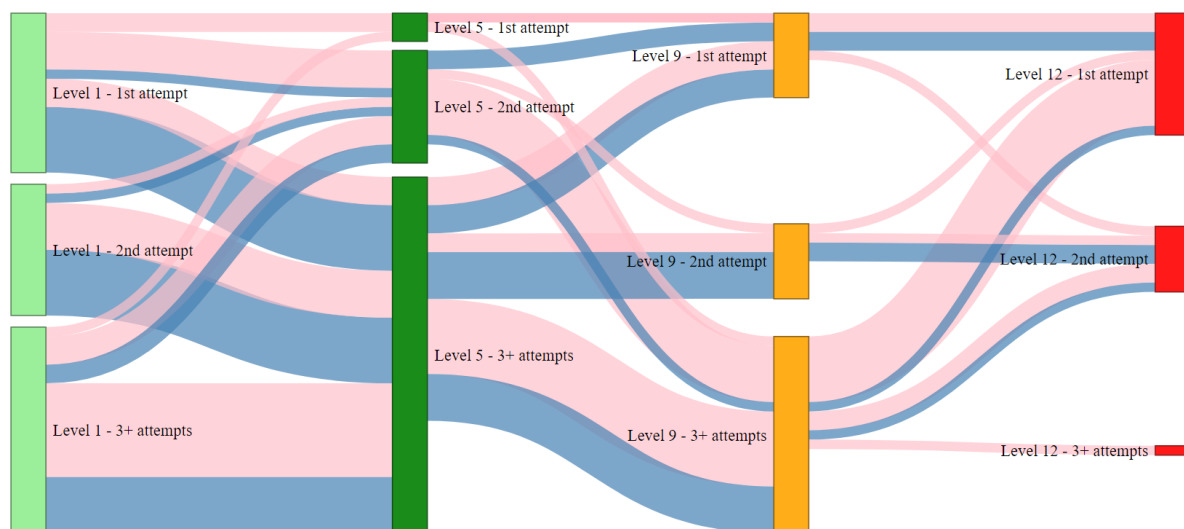


Figure 1: Sankey diagram illustrating the game flow and level plateaus based on the number of replays in 2018 (pink ribbon) and 2019 (blue ribbon). The number of attempts for each level equals to the number of replays minus 1 (i.e., 1st attempt represents 0 replays of that level). Levels 1, 5, and 9 were chosen as thresholds since they were found to be most troublesome for students while level 12 is the final game level.

At 5% significance level, there was no statistically significant difference in the overall game outcomes between 2018 and 2019 learning designs in regards to the number of replays ($W = 31172$, $p = 0.828$, $A = 0.495$), suggesting that the changes in the design did not increase the students’ learning outcomes.

However, there are notable differences with respect to which levels players got stuck between 2018 and 2019. Figure 1 illustrates that for both years, as the game level increased, the number of students decreased. Only a few students finished the game while many students ended their gameplay at level 9. It also shows that in both years, students struggled with level 1 despite support and guidance being offered from teachers in 2019. Many students struggled with level 5 as well. In both 2018 and 2019, approximately 33% of students had to replay level 5 two or more times. Nonetheless, in 2018, three students completed level 5 at the first attempt while in 2019, all of the students had to replay the level at least once. There was a 24% and 28% decrease in the number of students advancing from level 5 to level 6 in 2018 and 2019, respectively. Level 9, the framework of which is similar to that of level 5, was found to be problematic. However, there was a significant difference in the average number of level 9 replays between 2018 students and 2019 students with a moderate effect ($W = 242, p = 0.05, A = 0.688$). In both years, only a few students successfully completed the level at the first attempt reaching level 10. However, compared to 2018, in 2019 more players were successful in passing level 9 on their second attempt. In particular, in 2018, only 14% of students had to replay level 9 once compared to 31% of students in 2019. Nevertheless, only a few students advanced to level 10 and finished the game. The results indicate that targeted instruction as a tool to improve success in a math game had limited impact, although likely aided students to tackle advanced levels more successfully. Yet, most students got stuck on and ended their gameplay at level 9, which may be due to a better understanding of the game mechanic rather than an improved understanding of fractions. Further studies can be conducted in examining whether making levels easier will resolve the issue of level plateaus and increase the learning outcomes.

REFERENCES

- Alonso-Fernandez, C., Calvo, A., Freire, M., Martinez-Ortiz, I., & Fernandez-Manjon, B. (2017). Systematizing game learning analytics for serious games. *2017 IEEE Global Engineering Education Conference (EDUCON)*, 1111–1118. <https://doi.org/10.1109/EDUCON.2017.7942988>
- Conole, G. (2012). *Designing for learning in an open world*. Retrieved from <http://www.columbia.edu/cgi-bin/cul/resolve?clio10188861>
- Gunter, G. A., Kenny, R. F., & Vick, E. H. (2008). Taking Educational Games Seriously: Using the RETAIN Model to Design Endogenous Fantasy into Standalone Educational Games. *Educational Technology Research and Development*, 56(5/6), 511–537.
- Jayashanka, R., Hewagamage, K. P., & Hettiarachchi, E. (2018). Improving Blended Learning in Higher Education Through the Synergy between Learning Design and Learning Analytics. *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, 227–228. <https://doi.org/10.1109/ICALT.2018.00120>
- Nguyen, Q., Huptych, M., & Rienties, B. (2018). Linking Students' Timing of Engagement to Learning Design and Academic Performance. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, 141–150. <https://doi.org/10.1145/3170358.3170398>
- Peddycord-Liu, Z., Cody, C., Barnes, T., Lynch, C., & Rutherford, T. (2017). The antecedents of and associations with elective replay in an educational Game: Is replay worth it? *Proceedings of the 10th International Conference on Educational Data Mining* (pp. 40-47). Retrieved from http://educationaldatamining.org/EDM2017/proc_files/papers/paper_32.pdf

Development of a Learning Dashboard Prototype Supporting Meta-cognition for Students

Min Lu^{1*}, Li Chen¹, Yoshiko Goda², Atsushi Shimada¹, Masanori Yamada^{1**}

¹ Kyushu University, Japan; ² Kumamoto University, Japan

*lu@artsci.kyushu-u.ac.jp, ** mark@mark-lab.net

ABSTRACT: This poster presents an initial prototype design and development of a learning analytics dashboard supporting self-regulated learning, from which the students can benefit from LA more directly. The current stage of the development focuses on providing visualizations of learning processes and behaviors extracted from operation log data of an e-book system for self-monitoring. An overview of the reading paths and time of the slide pages in a class and a detailed view of the activities and learner-created content on the selected page are provided with a comparison of the class overall states and those of the learner. This work is expected to invoke the future developments and practical experiments of an LA dashboard supporting different phases of self-regulated learning.

Keywords: self-monitoring, dashboard, learning analytics, visualization

1 INTRODUCTION

Learning analytics (LA) with the large-scale educational log data obtained from e-learning environments can benefit both the instructors and learners with different kinds of feedback. Although researches of LA dashboards have become popular in recent years; however, as the most visible results are designed for the instructors, the learners cannot benefit from LA in a direct manner. On the other hand, monitoring the learner's own learning behaviors and processes is an important aspect of self-regulated learning because it helps learners to be aware of their weaknesses or deficiencies in their learning processes and regulates their learning strategies (Hofer et al., 1998). Visualization is useful for learners to be aware of what they have been doing and what they should do by making such information salient for them (Yen et al., 2018). Our prior research has designed a learning analytics dashboard supporting metacognition to improve self-regulated learning in online environments through the collection, analysis, and visualization of learning log data (Chen et al., 2019). This poster presents an initial prototype of the dashboard, focusing on supporting self-monitoring, which invokes future experiments, evaluations, and developments.

2 UI DESIGN OF THE LA DASHBOARD FOR STUDENTS

For self-monitoring, learners are expected to focus on the learning processes rather than the outcome only (Zimmerman, 1998). Thus, the dashboard intends to provide the students with the processes and behaviors visualized from learning log data in two types of views (Chen et al., 2019). The first one is an overview of the learning activities on all the slide pages in a class. The second one is the learning behaviors on a single page. As the point of self-evaluation is a comparison of one's own performance or behaviors based on certain criteria or standards (Belfiore & Hornyak, 1998), both views provide comparisons between the class's overall situations and the user's behaviors.

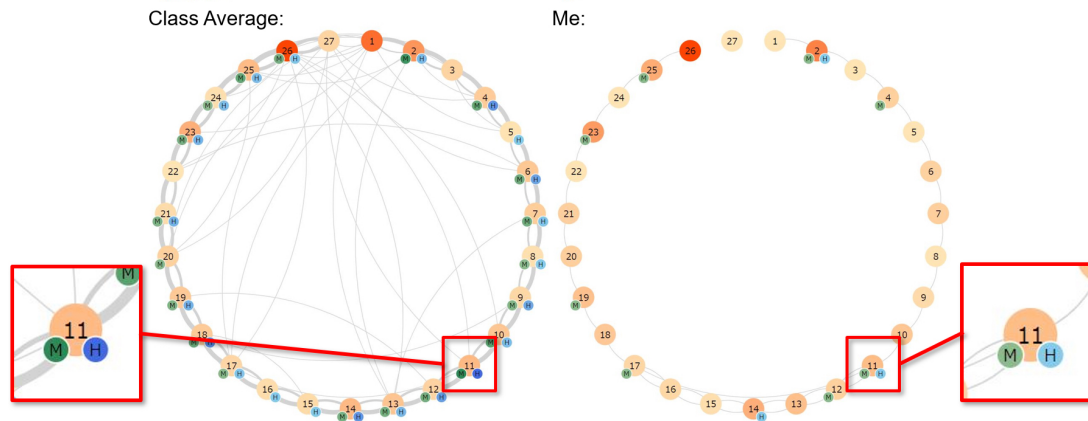


Figure 1: Overview of reading path and time on all the slide pages in a class

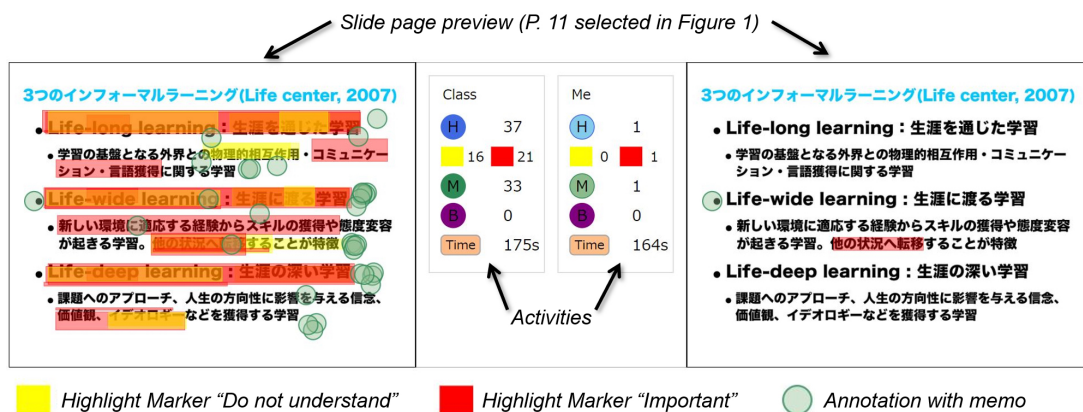


Figure 2: Detailed view of the activities and learner-created content on the selected page

2.1 Overview of All the Slide Pages in a Class

A graph to visualize the slide reading path and time with the nodes on a circle stand for pages and the links between the nodes stand for the reading path (Figure 1). The intensity of a node's color indicates the reading time spent on the page, and the thickness of a link shows the number of the page transit. The accessories, which are smaller circles attached to a page node, present the recorded learning behaviors, including highlight markers, memo annotations, on the page. The intensity of an accessory's color indicates the number of corresponded behaviors.

2.2 Detailed View of Each Slide Page

When the learner clicks a page node in the overview, the details of the reading time and learning behaviors will be displayed with the learner-created content overlapped on the slide page (Figure 2). The left view shows the average reading time of the class and all the highlight markers and memo annotations created on the page, while the right view shows those of the learner.

3 PROTOTYPE DEVELOPMENT

We mainly developed a data processing module and a web-based visualization module to realize the above design, mainly with the operation event logs from the e-book system of our university.

3.1 Data Processing Module

As the accumulated operation event logs are from different courses and students, this module at first filter the records according to slides and students' IDs, and then extract the sequences of page-transit related events by time after a data cleansing of events with too short intervals ($<0.5s$). From such sequences, the time spent on each page and the numbers of the "from-to" links between each pair of pages can be calculated. The overall states of the class can be obtained by summarizing the result from all the students. The results are stored in JSON files for the visualization module.

3.2 Web-based Visualization Module

This module is developed based on D3.js, and the visual elements (e.g., the nodes and links) are implemented as Scalable Vector Graphics (SVG), thus can be interactive to the user's mouse operation. This module also provides programming interfaces for setting up parameters of the visualization, such as the size of the visual elements, colors of the nodes, accessories, links, and intensity scales, and so on, for future developments.

4 PRELIMINARY EVALUATION AND FUTURE WORK

In the interviews with graduate students and teachers in our university, we obtained useful comments and suggestions. For example, the links should present the direction of page transits more clearly; the number displayed should be more meaningful to students; hyperlinks from the graph to other useful plugins of the LMS and e-book can be helpful; and so on. With the refined prototype, we plan to conduct formative experiments in the next step to clarify its effectiveness. In the future, the functions of the LA dashboard to support other phases of self-regulated learning, including knowledge monitoring, planning, and regulation, will be studied and developed.

ACKNOWLEDGMENT

This research is supported by a JST AIP Grant No. JPMJCR19U1, Japan.

REFERENCES

- Chen, L., Lu, M., Goda, Y., & Yamada, M. (2019). Design of learning analytics dashboard supporting metacognition. *Proceedings of 16th International Conference Cognition and Exploratory Learning in Digital Age (CELDA 2019)*, 175-182.
- Belfiore, P. J., & Hornyak, R. S. (1998). Operant theory and application to self-monitoring in adolescents. In: D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practice* (pp. 184-202). New York: Guilford.
- Hofer, B. K., Yu, S. L., & Pintrich, P. R. (1998). Teaching college students to be self-regulated learners. In: D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practice* (pp. 57-85). New York: Guilford.
- Zimmerman, B. J. (1998). Developing self-fulfilling cycles of academic regulation: An analysis of exemplary instructional models. In: D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practice* (pp. 1-19). New York: Guilford.
- Yen, M. H., Chen, S., Wang, C. Y., Chen, H. L., Hsu, Y. S., & Liu, T. C. (2018). A Framework for Self-Regulated Digital Learning (SRDL). *Journal of Computer Assisted Learning*, 34(5), 580-589.

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

From Text Mining to Evidence Team Learning in Cybersecurity Exercises

Author: Kaie Maennel

Tallinn University of Technology, Estonia
kaie.maennel@taltech.ee

Author: Joonsoo Kim

National Security Research Institute, South Korea
joonsoo@nsr.re.kr

Author: Stefan Sütterlin

Østfold University College, Norway
stefan.sutterlin@hiof.no

ABSTRACT: Team-based cybersecurity exercises are popular learning methods to reduce the skill gap in digital security. However, assessing team learning remains an unsolved research question. Our ongoing research focuses on unobtrusive team learning measurements in exercises using existing reporting mechanisms. We apply text mining techniques to the situational reports (SITREPs) in a series of large annual international training events with over 1,000 participants every year. We operationalize the information processing required for learning (share, store, retrieve) with metrics obtained from SITREPs, which we hypothesize to reflect team-level learning in the cyber domain. This poster focuses on our methodology and will present initial results. The obtained knowledge will further form an evidence-based foundation for designing technical solution for semi-automated, scalable and unbiased SITREPs evaluation and provide timely and comparable feedback to the teams.

Keywords: learning analytics, team learning, text mining, situational reports, cybersecurity exercises

1 INTRODUCTION AND BACKGROUND

The cybersecurity operational work often takes place in teams (e.g., incidence response teams) and requires effective knowledge sharing and collaboration between individuals, teams and organizations. Therefore, training events have a strong team learning component. When designing effective learning, the educators and organizers need to understand the dynamics of how teams learn and what are indicators of successful team learning. However, assessing team learning in a scalable way and avoiding invasive assessment methods (such as observation, testing), remains an unsolved research question in such learning environments.

Our ongoing research focuses on unobtrusive team learning measurement in the team-based cybersecurity exercises using existing reporting mechanisms (situational reports, SITREPs) (Maennel et al. 2019). We consider team learning as sharing, storage, and retrieval processes that are intertwined and need to take place for group learning to occur (Wilson et al. 2007). In SITREPs, teams are expected to demonstrate sensemaking in routine operational work or related to a specific

situation (Franke and Brynielsson 2014). Teams need to analyze the situation, and the main task is to understand that not all attacks/events have the same strategic impact (Doupe et al. 2011) with respect to the exercise objectives. Therefore, we consider SITREPs as collective repositories and expression of team knowledge. We apply text mining techniques to analyze situational reports recorded at a series of large international cybersecurity exercises, with over 20 national or multinational teams, over a time-period of four years.

As a first step, we aim to develop baseline for an ontology for cybersecurity vocabulary (situational reports, chats, etc.), as the terminology and expressions differ from mainstream natural language processing and such dictionary is lacking in cybersecurity domain. While prior work has introduced some concepts, our contribution is a representation of words/concepts and their relationships in the cybersecurity and team learning context. Our work can be openly accessed, applied and connected to other sources of information (such as automated scoring, human feedback, and red teaming).

2 METHODOLOGY

We hypothesize that there are general characteristics of SITREPs, that show that teams are sharing, storing and retrieving the information. Over time this shows the progress of teams and thus evidences how teams are learning. The learning indicators that have been identified based on initial qualitative review of SITREPs are discussed below:

- Number and key words of events described to ensure are incidents identified and reported by the team are in line with the exercise scenario (learning objectives);
- Length of SITREPs or specific section, e.g., longer vs. shorter executive summary;
- Vocabulary used by the teams to assess:
 - Sentiment and learner engagement, e.g., use of words such as "gamenet";
 - Detailedness and level of understanding of the situation/event occurred, e.g., use of adjectives to describe an attack;
 - Cognitive focus measure by applying Hybrid Space concept (tactical/strategic and cyber/physical dimensions) (Jøsok et al. 2016) to identify possible statements referring to the degree of awareness, performance, uncertainty, quality of decisions made, etc. See Figure 1 for an example of words used and grouped in HS concept.

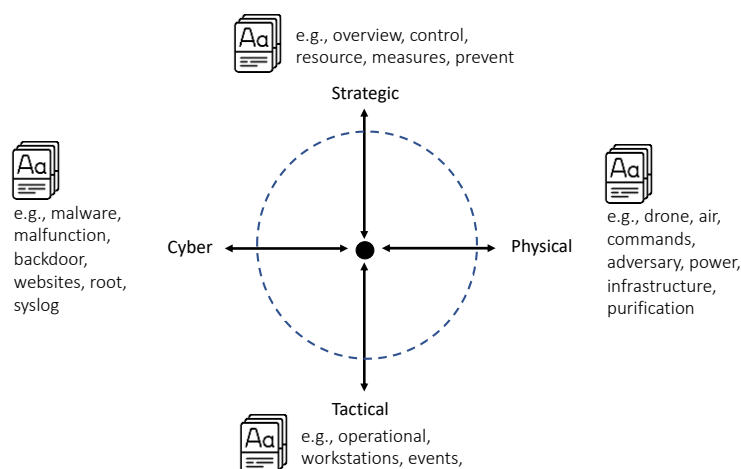


Figure 1: Example of SITREPs vocabulary mapped to Hybrid Space (Jøsok et al. 2013)

These learning indicators are correlated to other information such as:

- Team size or composition, e.g., whether larger teams are facing more challenging to share, store and retrieve what they have learned;
- Survey feedback with the team's self-assessment of learning achieved at individual, sub-team(s) and team level;
- Overall performance of the teams and other information collected during the exercises.

We use NVivo¹ for qualitative analysis involving labelling and coding the data to recognize similarities and differences (code, i.e. key words/expressions counting), and Python and LightSide² for text mining and correlation analysis.

3 CONCLUSION AND FUTURE WORK

We have carried out an initial analysis. For example, our findings indicate that reflections about cyber-related real-world consequences and teams' critical assessment of their control over situation may be a good indicator for metacognitive processes and learning. Namely, the use of words such as "control" or "out-of-control" are stronger associated with the physical rather than the cyber domain (Maennel et al. 2019).

We operationalize the information processing required for learning with metrics and correlations obtained from SITREPS, which reflect team-level learning in the cyber domain. Further, we validate that inter-correlations (e.g., "semantic proximity" (Slimani 2013)) are suitable learning indicators to measure learning in such cybersecurity exercises. The developed ontology and data analysis results will form an evidence-based foundation for designing technical solution for semi-automated, scalable and unbiased evaluation and provide timely and comparable feedback to the teams.

REFERENCES

- Doupé, A., Egele, M., Caillat, B., Stringhini, G., Yakin, G., Zand, A., ... & Vigna, G. (2011, December). Hit'em where it hurts: a live security exercise on cyber situational awareness. In *Proceedings of the 27th Annual Computer Security Applications Conference* (pp. 51-61). ACM.
- Franke, U., & Brynielsson, J. (2014). Cyber situational awareness—a systematic review of the literature. *Computers & Security*, 46, (pp. 18-31).
- Jøsok, Ø., Knox, B. J., Helkala, K. M., Lugo, R., Sütterlin, S., Ward, P. (2016). Exploring the Hybrid Space: Theoretical Framework Applying Cognitive Science in Military Cyberspace Operations. *Lecture Notes in Computer Science*, 9744, (pp. 178-188).
- Maennel, K., Kim, J., & Sütterlin, S. (2019, June). Team learning in cybersecurity exercises. In *5th Interdisciplinary Cyber Research Conference* (p. 17).
- Slimani, T. (2013). Description and evaluation of semantic similarity measures approaches. arXiv preprint arXiv:1310.8059.
- Wilson, J. M., Goodman, P. S., & Cronin, M. A. (2007). Group learning. *Academy of Management Review*, 32(4), (pp. 1041-1059).

¹ <https://www.qsrinternational.com/nvivo/nvivo-products>

² <http://ankara.lti.cs.cmu.edu/side/>

LA Platform in Junior High School: Trends of Usage and Student Performance

Rwitajit Majumdar, Hiroyuki Kuromiya, Kiriko Komura, Brendan Flanagan, & Hiroaki Ogata
Kyoto University
dr.rwito@gmail.com

ABSTRACT: This study reports the initial trends found in the pilot phase of a Learning analytics (LA) platform adoption at a Junior high school in Japan. The LA platform includes a Learning Management System (LMS), e-Book reader, and analytics dashboard that is accessible to both teachers and students. The interaction logs of those learning tools and mid-term test score for the third-year junior high school mathematics class with 120 students were analyzed. The result highlighted that a group of students who voluntarily explored the dashboard performed significantly better than the group of students who did not check the dashboards. However, both the groups' e-Book interaction counts were not significantly different. This initial result was encouraging as the evidence was extracted from the data collected without any specific interventions. The findings also motivated further investigation in the usage pattern of the LA platform and design of interventions.

Keywords: Evidence-based Education, TEEL Platform, LAViEW, BookRoll

1 BACKGROUND

Previous works in LAK considering learning analytics in schools, focused on personalization of intelligent tutoring system for middle school mathematics class (Fancsali & Ritter, 2014), predicting failures early in course (Jiménez-Gómez et al., 2015), or at-risk students with respect to graduating high school (Aguar et al., 2015). However, studying trends and collecting evidences from the wild in any face-to-face teaching-learning context at school level is rarely discussed. In this paper we report the initial trends of usage of an implemented Learning Analytics (LA) platform during its pilot phase. Additionally, the study reports relationship between students performance and learning activities.

1.1 Technology Context – Learning Evidence Analytics Platform

The LA platform includes Moodle as LMS, BookRoll, an eBook reader, and LAViEW, as LA dashboard (Majumdar et al., 2019). Both teachers and students can access the learning tools. This integrated LA system was introduced at a junior high school level. Teachers were guided to use the eBook reader features to design learning activities, and then orchestrate them in their class. LAViEW visualized the reading interaction logs. Teachers could review their students learning behaviors. Their students (Learners) could also reflect on their own learning data.

The authors highlight three features of LAViEW that helps teachers in school and LA researchers at universities to collaborate. The first feature is a function to upload offline test scores. As many of the formative assessments in Japanese school were traditionally paper based. The function allowed teachers to upload the tabulated test scores and stored it directly in the Learning Record Store (LRS). Like other learning logs, these records too were linked only to the system generated pseudonymized

ID of the learner. The performance data is then accessible to the researchers for developing learner models along with the other learning logs in the LRS. The second feature was an implementation of xAPI statement-based logging of LA dashboard interactions. It logged which visualized indicators (graphs) the user checked LAViEW. To our knowledge, such standardized logging of dashboard action is not available in previous literature. It would assist tracking monitoring and reflection behaviors of users. The third feature is an evidence portal which assists evidence extraction from the collected data stored in the LRS. Extracting such as learning logs, test scores, and dashboard access logs. The portal's objective is to systematically collect evidence within the teaching-learning system itself.

2 RESEARCH CONTEXT AND FINDINGS

The LA platform was introduced as a pilot across three schools giving access to students in both junior high and high schools. For initial analysis, this study considered only one junior high school, and focused on mathematics course for the third-year students at junior high school in urban school district in Japan. During the exploratory data analysis phase of the logs from the initial period, the researchers were motivated to examine *“What are the trend of using the LA platform and its relationship to the student's performance?”*

2.1 Descriptive Statistics of Overall Usage and Collected Data

This study focused on math class taught for the third-year junior high school students. A total of 120 students were enrolled in the class composed of three sections on Moodle LMS with 40 participants each. All the three Moodle courses and the reading resources shared in BookRoll, such as text books, teacher's notes, and practice quizzes were the same for all of these three sections. Moreover, LAViEW, the dashboard linked to the Moodle, displayed information regarding students reading behaviors and Moodle activities. The teacher could review aggregated or individual student information while students could reflect on their own learning data. For the current study, course interaction logs related to textbook reading and LA Dashboard viewing were considered during the period of June to September 2019. The reading log is only of the students (N=120) and the Dashboard log includes the teacher (N=121). A total of 6,639 reading logs over 53 unique days were related to only opening and annotation in the textbook. Similarly, 541 logs related to specific graph seen in LAViEW dashboard was collected over 30 unique days of usage. Apart from the interaction log data, we also collected students' mid-term test score in the beginning of June 2019. The teachers uploaded the test scores through the LAViEW dashboard.

2.2 Initial Findings: Dashboard Usage and Relation to Performance and Reading

Though students were not instructed to check the dashboard for any specific purpose, an interesting finding from the analysis of the log data showed that some students (n=16) still accessed LAViEW from the course Moodle. This observation further motivated to compare the test performance and reading activities for the group of students who choose to review the dashboard (In, n=16) vs. one did not (Out, n=104). Figure 1. presents the count of particular graphs viewed by the teacher and all the students. While the teacher checked the detailed graphs related to the annotations of the students (overlay of the annotations in contents review, memos and detail in data table) more, the students focused on the overview summary, their reading time and completion information.

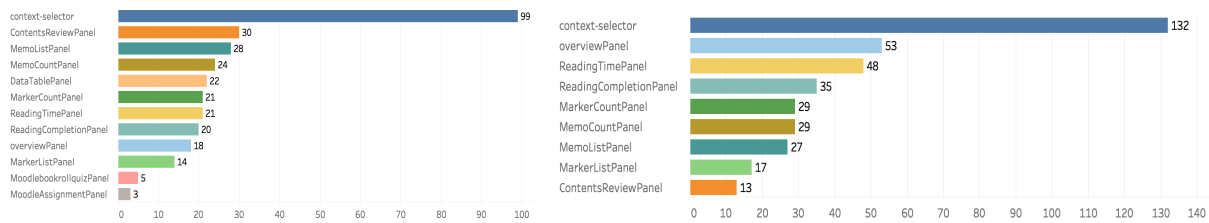


Figure 1: LA Dashboard Graphs Accessed. a. Teacher b. Students (n=16)

The study found there were significant differences in the test score (T-test). However, there was no difference in the count of reading logs for those two groups (Mann–Whitney U test). The group of students who checked the dashboard (In, n=16) had significantly higher ($p=0.004$) test score (Mean=71.56, Std dev.=20.92) than the ones (Out, n=104) who did not see dashboard (Mean=59.08, Std dev. = 14.96). The number of reading logs were not significantly different ($p=0.093$). Figure 2 shows the distribution of the test score and number of reading logs for both groups.

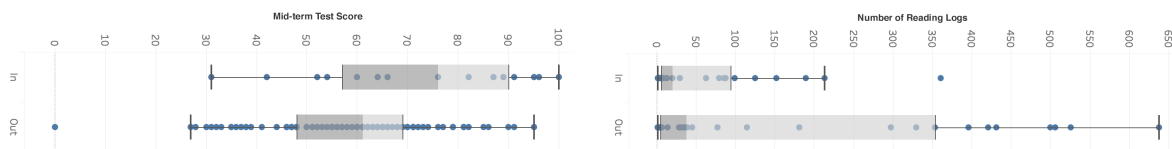


Figure 2: Distribution of attributes of two groups a. Mid-term Test Score b. Reading Logs

This initial analysis uses interaction data collected in regular classroom without any specific teaching-learning interventions. Hence it motivates further examination in the natural usage pattern of the components of the LA platform and then co-design learning interventions involving the teachers and LA researchers to meaningfully integrate the dashboard in the students learning experience. Future studies are planned to implement and evaluate such interventions.

ACKNOWLEDGEMENT

This work was partly supported by NEDO Special Innovation Program on AI & Big Data 18102059-0, JSPS KAKENHI Grant-in-Aid for Scientific Research (S) Grant Number 16H06304 and JSPS KAKENHI Research Activity Start-up Grant Number 18H05746.

REFERENCES

- Fancsali, S. E., & Ritter, S. (2014). Context personalization, preferences, and performance in an intelligent tutoring system for middle school mathematics. In Proc. of 4th LAK (pp. 73-77).
- Jiménez-Gómez, M. Á., Luna, J. M., Romero, C., & Ventura, S. (2015). Discovering clues to avoid middle school failure at early stages. In Proc. of 5th LAK (pp. 300-304).
- Aguiar, E., Lakkaraju, H., Bhanpuri, N., Miller, D., Yuhas, B., & Addison, K. L. (2015). Who, when, and why: a machine learning approach to prioritizing students at risk of not graduating high school on time. In Proc. of 5th LAK (pp. 93-102).
- Majumdar, R., Akçapınar, A., Akçapınar, G., Flanagan, B., & Ogata, H. (2019). LAView: Learning Analytics Dashboard Towards Evidence-based Education. In Companion Proc. of 9th LAK.

Analysis of Students' Self-Regulatory Strategies in MOOCs

Jorge Maldonado-Mahauad

Universidad de Cuenca
Pontificia Universidad Católica de Chile
jorge.maldonado@ucuenca.edu.ec

Mar Pérez-Sanagustín

Universite Paul Sabatier Toulouse III
Pontificia Universidad Católica de Chile
mar.perez-sanagustin@irit.fr

ABSTRACT: This poster presents a methodological approach to investigate the learning strategies that students use in online learning environments. Using mixed methods as a baseline, this methodological approach proposes combining students' digital traces with self-reported data for identifying SRL strategies. Specifically, the methodological approach uses process-mining techniques for automatically detecting students' strategies from learners' trace data and combines it with the students' SRL profiles extracted from self-reported data to explain these strategies. This method, which has been applied in MOOCs, sets the basis for the study of SRL as a process in any digital learning environment, and opens up the debate for new research avenues for personalization and adaptation.

Keywords: Self-regulated Learning, Massive Open Online Courses, Process Mining, Learning Outcomes

1 INTRODUCTION

Studies point out that self-regulation is a crucial higher-order skill required to adapt to the continually changing professional environments of the 21st century (Häkkinen et al., 2017). Self-regulated learners use cognitive, metacognitive and resource management strategies to plan, manage and control their learning process to achieve their goals and persevere until they succeed. In the current literature, self-regulated learning has been mostly studied on the basis of different theoretical models (Panadero, 2017). However, one of the main difficulties when studying SRL strategies is operationalizing these models for analyzing how SRL occurs (Jakesov & Kalenda, 2015). In the past years, the data collected by digital learning environments and the methodologies proposed by the Learning Analytics community open up new possibilities for understanding SRL strategies from learners' actual behavior. Depending on the model that is assumed, self-regulation can be studied from two perspectives: (1) as an aptitude or (2) as an event (Kizilcec et al., 2017). From the aptitude perspective, many instruments have been developed in the last decade to measure the students' SRL strategies profile in online environments, being questionnaires one of the most commonly used (Kizilcec et al., 2017). From the event perspective, researchers have started to study SRL by analyzing students' trace data collected by the digital environments during their study sessions (Jakešov & Kalenda, 2015). Our methodological approach follows both the aptitude approach and the event approach by combining data from self-reported questionnaires with data from students' traces.

2 METHOD

We propose using process-mining techniques to automatically detect the most common strategies and then analyze these strategies according to learners' SRL profiles detected from the students' self-reported questionnaires. As an example of how this methodological approach is applied, we present a case study with data collected from 3 MOOCs in Coursera.

2.1 Context: Course and Sample

A sample of N= 3,458 online learners in three different MOOCs on Coursera were considered (n = 2,035 in engineering, n = 497 in education and n=926 in management). The average age was 32.0 (SD. 11.07), one quarter of learners were women and 88% held a bachelor's degree or higher (14% a master's or Ph.D.). Data collection occurred between April and December 2015.

2.2 Measuring SRL strategies: as an aptitude and as an event

To assess SRL-strategies as an aptitude, students completed an SRL questionnaire with 24 statements related to six SRL strategies that was adapted from multiple established instruments (Maldonado-Mahauad et al., 2018). The six SRL strategies assessed are goal setting, strategic planning, self-evaluation, task strategies, elaboration and help-seeking. The SRL measure exhibited high reliability for all strategy subscales with a Cronbach's alpha of at least 0.70.

To assess SRL-strategies as an event, we adapted the Process Mining PM² method by Van Eck, Lu, Leemans, & Van Der Aalst (2015). The PM² method is structured into four stages: **(1) Stage 1-extraction** - the data is extracted from the Information System Data Bases (Coursera in our case), **(2) Stage 2-event log generation** - a table with valuable information is defined from trace data for generating an event log that includes the concepts of "*case*" (execution of a process), "*activities*" (steps of the process), and "*temporal order of the activities*" (timestamp of the activities), **(3) Stage 3-model discovery** - process mining discovery algorithms are applied to the event log in order to automatically mine a process model describing the observed behavior of the process, and **(4) Stage 4-model analysis** - the discovered process models are analyzed in order to understand the observed behavior.

3 RESULTS

3.1 SRL strategies in MOOCs

Three types of learners were identified using this method (see Figure 1): **(1) sampling learners**, who have a low activity in the course, just watch a single video-lecture or start "sample" at the beginning of the course exploring materials with the course already started); **(2) comprehensive learners**, who can be considered as more self-regulated, because they developed a variety of learning strategies per session, watching more video-lectures on average per session than other learners. They tend to follow the path that is provided by the course structure. They also invest more time watching video-lectures); and **(3) targeting learners**, who have similar SRL scores as comprehensive learners, but targeting learners are more strategic and focus their efforts on assessments to achieve performance-oriented objectives and exhibit less engagement overall).

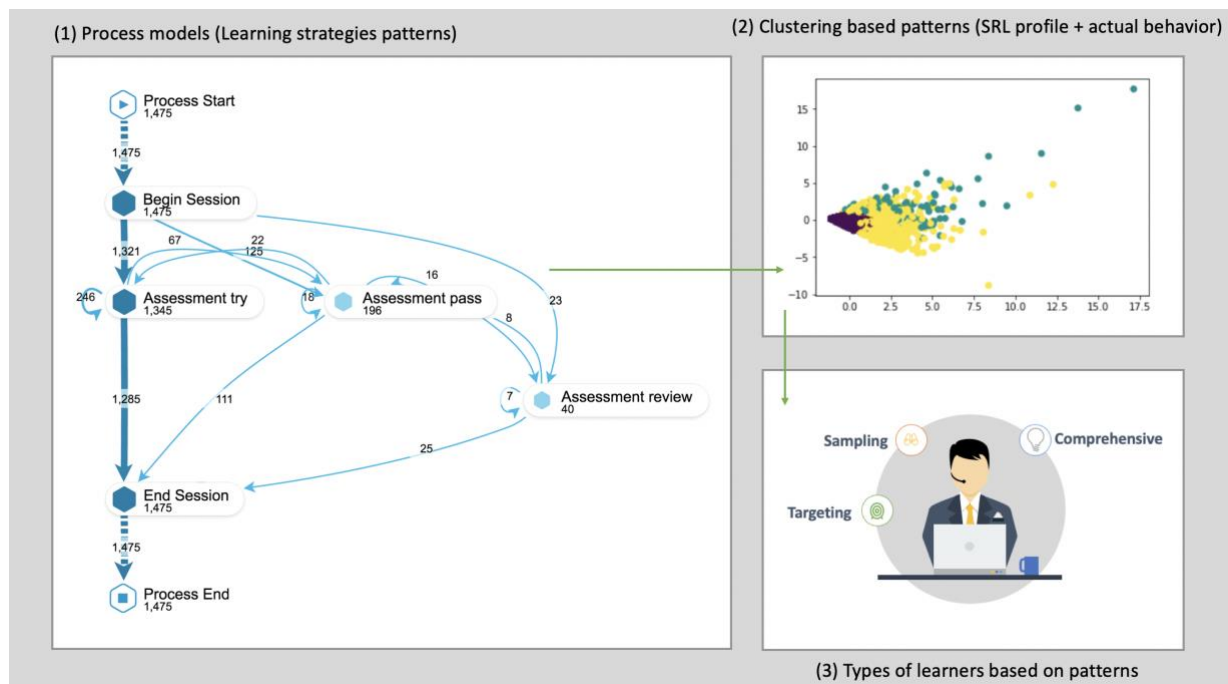


Figure 1: Learners classified by their learning strategies patterns (aptitude and event based)

4 CONCLUSIONS

This poster presents a method for investigating SRL strategies in MOOCs that combine an aptitude-based approach with an event-based approach. The results of an empirical study show that this method can be useful to identify learners' patterns in MOOCs, using the combination of the two approaches instead of only one of them (this differs from previous studies).

5 ACKNOWLEDGEMENTS

Work funded by Dirección de Investigación de la Universidad de Cuenca (DIUC), under Analítica del aprendizaje para el estudio de estrategias de aprendizaje autorregulado en un contexto de aprendizaje híbrido (DIUC_XVIII_2019_54) and the LALA project (grant no. 586120-EPP-1-2017-1-ES-EPPKA2-CBHE-JP).

REFERENCES

- Häkkinen, et al. (2017). Preparing teacher-students for twenty-first-century learning practices (PREP 21): a framework for enhancing collaborative problem-solving and strategic learning skills. *Teachers and Teaching*, 23(1), 25-41.
- Jakešov., J., & Kalenda, J. (2015). Self-regulated learning: Critical-realistic conceptualization. *Procedia-Social and Behavioral Sciences*, 171, 178–189.
- Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & education*, 104, 18-33.
- Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R. F., Morales, N., & Munoz-Gama, J. (2018). Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. *Computers in Human Behavior*, 80, 179-196.
- Van Eck, M., Lu, X., Leemans, S., & van der Aalst, W. (2015). PM²: A Process Mining Project Methodology. In *Advanced Information Systems Engineering* (pp. 297–313).

Implementing learning analytics in schools: Towards a theory-based and data-driven framework for analytics implementation

Claudia Mazziotti

Technische Universität München

claudia.mazziotti@tum.de

Vitomir Kovanović, Shane Dawson & George Siemens

University of South Australia

vitomir.kovanovic@unisa.edu.au; shane.dawson@unisa.edu.au;

george.siemens@unisa.edu.au

ABSTRACT: While learning analytics have been well investigated in higher education, little is known about how its outcomes and approaches can be transitioned to other educational contexts such as secondary schools. This study identifies the conditions as well as the limiting factors influencing learning analytics uptake within schools. The goal of this research is to develop a theory-based and data-driven framework for guiding learning analytics implementation processes. This poster focuses on the data-driven part of the envisioned framework; that is, conducting interviews with school practitioners and deriving implications from the interviews for the framework development. To identify a variety of challenges and enabling conditions for learning analytics in schools, we conducted (and are currently conducting) interviews with practitioners from Australian secondary schools. To incorporate a broad range of perspectives and capture different levels within schools, our target group includes both administrative leaders and teachers. Building upon previous work and a top-down and bottom-up approach, we iteratively developed and improved an interview protocol and a coding scheme. Here we present preliminary findings and outline the future analysis steps.

Keywords: Learning analytics school adoption, learning analytics in secondary schools, learning analytics adoption framework

1 TOWARDS LEARNING ANALYTICS IN SCHOOL

Over the past decade, there has been a growing body of evidence demonstrating how the use of learning analytics (LA) facilitated the work of institutional leaders and teachers and also supported productive student learning (Joksimović, Kovanović, & Dawson, 2019). By monitoring different forms of learning and by combining different sources of data about student learning, sophisticated statistical models help(ed) to predict students' performance and to design educational interventions accordingly. Most of the work in the field of LA, however, has focused on and been implemented in tertiary education (Tsai et al., 2018). The comparably fewer LA systems that were specifically designed for the implementation in secondary education have only been adopted slowly even though they delivered promising results (Lodge, Horvath, & Corrin, 2018). The goal of this study is to draw research attention towards the implementation of LA in school by developing a theory-based and data-driven framework guiding the future LA implementation process. More precisely, we pose the question of what are challenges and affordances for the implementation of LA in school.

To address our research question, we followed a similar approach like Colvin and colleagues (2016) which examined the use and adoption of LA in higher education contexts. In the first, *theory-based step*, we reviewed existing models about i) how to implement LA - our target educational innovation - into other formal learning contexts (i.e., universities) and ii) existing approaches about how to

implement other educational innovations into school, which is our target learning context. In a second *data-driven* step, we conducted and are still conducting interviews with school practitioners and posed questions around the perception, potential and challenges of LA in school. This way, we are able to examine the specific challenges and affordances for the implementation of our target educational innovation in our target formal learning context. The results from both steps will be fed into the aforementioned framework. The focus of the work presented here lies in the setup and preliminary results of our interviews (i.e. the second step).

2 METHODOLOGY

2.1 Sample and data collection

To examine what are the enabling conditions and limiting factors for implementing LAs in secondary schools, we are currently conducting semi-structured interviews with school practitioners from the three branches of Australian secondary schools (i.e., independent private schools, catholic schools and public schools). First, we obtained ethics clearances from the administering university as well as from the Department of Education, Catholic Education Association, and independent schools. To identify challenges and affordances on all school levels (i.e., individual, class, school, community), we are interviewing at least one senior administrative leader and one or two teachers in each school. The administrative leaders include school principals, assistant principals, or technology coordinators who are involved in decision making around technology and LA matters. Similarly, teachers are primarily from STEM areas as they typically adopt more educational technologies and tend to have a good understanding of analytical and statistical issues that are the foundation of LAs. We plan to recruit 15–18 school practitioners from Australian schools. At present, we already conducted ten interviews, each lasting between 40–85 minutes. Before the interviews, we received consent from all participants and provided them with the participant information sheet.

2.2 Interview protocol

To investigate challenges and affordances of implementing LA in school through semi-structured interviews, we developed an interview protocol by building upon interview protocols for examining tertiary LA adoption in university and educational technology adoption in school (e.g., Colvin et al., 2016). Both protocols were based on theories of educational leadership, institutional change, and innovation adoption. The combined interview protocol consisted of 20 open-ended questions, covering adoption challenges, frequency of use, teacher support, and the relation between LA and educational standards.

2.3 Coding manual

To develop a coding manual, we identified key themes by following top-down and bottom-up approaches. In the top-down approach, we grounded our themes and codes in the theories described in Section 2.2. In the bottom-up approach, we iteratively reviewed the interview transcripts and identified themes that were not yet covered by the theory-driven codes. Overall, our coding manual covers eleven broader themes, each with three to five codes. The themes focused on educational technologies and LA adoption, such as *frequency of use*, *culture*, *type*, *purpose*, *scope*, *goals*, *approach*, *sources of data*, *school infrastructure challenges*, *professional support*, and *practitioners' concerns*. To improve interrater reliability of the coding process, we selected anchor examples for the initial calibration of the coding procedures.

3 PRELIMINARY INSIGHTS AND FUTURE WORK

To examine the specific challenges and affordances for the implementation of LA within the school context, we focused on the following more fine-grained research questions:

- RQ 1:** How are educational technologies currently used in schools?
- RQ 2:** How are LA perceived within the school context?
- RQ 3:** What is from a practitioner's point of view the added value of LA?
- RQ 4:** What are the challenges of implementing LA – and educational technologies more broadly – within secondary school contexts?

Even though interviews are still being conducted and transcript coding is not completed, we can already deliver some preliminary results and report about anecdotal evidence concerning each research question. In terms of RQ1, school practitioners reported using educational technologies such as learning management systems, mostly on a daily basis. With regards to RQ2, school practitioners' perceptions ranged from not being aware of the existence of LA to mapping out an entire LA project at their school. Concerning RQ3, if practitioners are aware of LA, then the added value from their perspective lies in diagnosing problems on different school levels and synthesising the different sources of online data and offline data. Finally, in terms of RQ4, practitioners were concerned about ethics and privacy issues as well as the extent to which they receive formalised support throughout the entire implementation process. In this context, the next step will focus on reviewing the interviews and finalising the appropriate indicators and codes. In the last step, we will identify connections between codes (e.g., how do different challenges – such as ethics and practitioners' data literacy – relate to each other) and draw a coherent network between them. For this, we will apply a widely used Epistemic Network Analysis (Shaffer, Collier, & Ruis, 2016) technique. Specifically, the latter, in combination with implications from theory (cf. step 1), will help us to develop a unified framework guiding the LA implementation process in schools.

REFERENCES

- Colvin, C., Rogers, T., Wade, A., Dawson, S., Gasevic, D., Buckingham Shum, S., ... Fisher, J. (2016). *Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement* [Report]. Retrieved from Australian Government Office for Learning and Teaching website: <https://opus.lib.uts.edu.au/handle/10453/117173>
- Joksimović, S., Kovanović, V., & Dawson, S. (2019). The Journey of Learning Analytics. *HERDSA Review of Higher Education*, 6, 37–63. <http://www.herdsa.org.au/herdsa-review-higher-education-vol-6/37-63>
- Lodge, J. M., Horvath, J. C., & Corrin, L. (Eds.). (2018). *Learning Analytics in the Classroom: Translating Learning Analytics Research for Teachers*. Abingdon, Oxon; New York, NY: Routledge.
- Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A Tutorial on Epistemic Network Analysis: Analyzing the Structure of Connections in Cognitive, Social, and Interaction Data. *Journal of Learning Analytics*, 3(3), 9–45. <https://doi.org/10.18608/jla.2016.33.3>
- Tsai, Y.-S., Moreno-Marcos, P. M., Jivet, I., Scheffel, M., Tammets, K., Kollom, K., & Gašević, D. (2018). The SHEILA Framework: Informing Institutional Strategies and Policy Processes of Learning Analytics. *Journal of Learning Analytics*, 5(3), 5–20–25–20. <https://doi.org/10.18608/jla.2018.53.2>

Analytics of Multimodal Learning Logs for Page Difficulty Estimation

Tsubasa Minematsu*, Atsushi Shimada, and Rin-ichiro Taniguchi

Kyushu University, Japan

*minematsu@ait.kyushu-u.ac.jp

ABSTRACT: We analyzed different types of learning activity logs to estimate where on pages students had difficulty during learning. Logs for this study were collected from an e-book system and an eye tracker. The e-book system recorded students' highlights, and the eye tracker collected eye movement data while students read. These data enabled us to estimate page difficulty. Our experiment compared these results to investigate data combination as a possibility for page difficulty estimation. We then obtained findings about the similarities between these results and discussed performance improvement through data combination.

Keywords: multimodal data, eye movement, machine learning, neural network

1 INTRODUCTION

Recently, various kinds of learning activity logs have been used to understand students' learning behaviors based on machine learning methods. Digital learning systems can collect a large number of student-system interactions, for instance, from more than 100 students simultaneously (Ogata et al. 2015). However, because logs are stored when students perform operations only, it is difficult to collect more detailed learning logs, such as those that capture reading behaviors on e-book pages. Eye trackers are used for measuring detailed student activity such as eye movements (Minematsu 2019). However, collecting data from many students is time-consuming. We will propose that multimodal data—or combining a large volume of learning logs that have been collected by digital learning systems with a small amount of eye movement data—can be helpful in compensating for such disadvantages. In addition, multimodal data can promote machine learning-based approaches in learning analytics. This study has focused on page difficulty estimation in teaching material for investigating the possibility of data combination. The investigation relied on learning activity logs that were recorded by an e-book system and students' eye movement data from their readings of a digital textbook. We showed page difficulty estimation analysis for each set of learning activity data as preliminary experimental results.

2 METHOD

2.1 Data Collection

We collected two types of student learning logs, namely student eye movement data as reading logs and highlights from students who self-studied or participated in an on-site lecture. Firstly, eye movement data were collected from 15 undergraduate students using an eye tracker. The 15 students self-studied a statistical test and correlation using an e-book system. In a dark room, we measured their eye movements individually. We also asked the 15 students to give their subjective impressions

of various pages' difficulty. Secondly, highlights were collected from approximately 1,200 students using the same e-book system and teaching materials that were utilized when the eye movement data were collected. A teacher delivered lectures in some classes, and all 1,200 students learned the same content. The students had the opportunity to add or delete highlights on the e-book's text and figures when they found difficulty. Highlights were added by 81 students.

2.2 Analyzing Students' Eye Movements

In order to identify where students had difficulty, we investigated the relationship between students' eye movements and their subjective impressions of page difficulty. The eye movement data we collected were utilized in the context of the neural network-based method proposed in Minematsu's "Region-wise page difficulty analysis using eye movements" (2019). This method generated relevance maps that represented where students found difficulty on the page. Firstly, the relationship was modeled after the neural network; that is, page difficulty was inferred based on corresponding eye movement data. Next, the author used layer-wise relevance propagation (LRP) (Bach et al. 2015) to analyze the inference's most relevant student eye movement data. LRP enabled the visualization of the specific eye movement(s) responsible for contributing to the neural network's decision, thus enabling us to determine each page's relevance map. In this study, we binarized the relevance maps to get binary masks. In addition, we analyzed a distribution of gaze points for each page in order to understand where students were looking. The distribution is called a gaze map, and gaze maps were generated by following the same procedure as described in Minematsu's article (2019). We also binarized the gaze maps.

2.3 Analyzing Students' Highlights During an On-Site Lecture

We made highlight maps of each page to investigate where students found difficulty. The highlight maps represented the number of highlights students added at the same location on each page. In order to focus on difficult content areas, we extracted highlighted regions with more than half the total number of highlights on each page. As a result, this procedure also generated binarized highlight maps.

3 RESULTS AND DISCUSSION

We compared the highlight maps with two maps that were generated from eye movement data. Figure 1 shows the highlight, relevance, and gaze maps for the three most difficult pages. These three pages were chosen based on the 15 students' subjective impressions of page difficulty that were used in the analysis of eye movements. We observed that, according to the gaze maps, the students read the pages uniformly. The relevance maps extracted a part of the gaze maps, and parts of the relevance maps corresponded to some parts of the highlight maps. We quantitatively evaluated the similarities between the maps that were generated from the different types of the data. Table 1 shows the precision, recall, and F-measure values that were calculated by comparing the highlight maps with the two maps that were generated from eye movement data. According to Table 1, the relevance maps achieved higher precision than the gaze maps in the top three most difficult pages. This indicates that some areas in the relevance maps more accurately corresponded to some areas in the highlight maps. The recall values were the highest in the gaze maps because the students read the content of each page uniformly. These results indicate that the highlight and eye movement data contain information

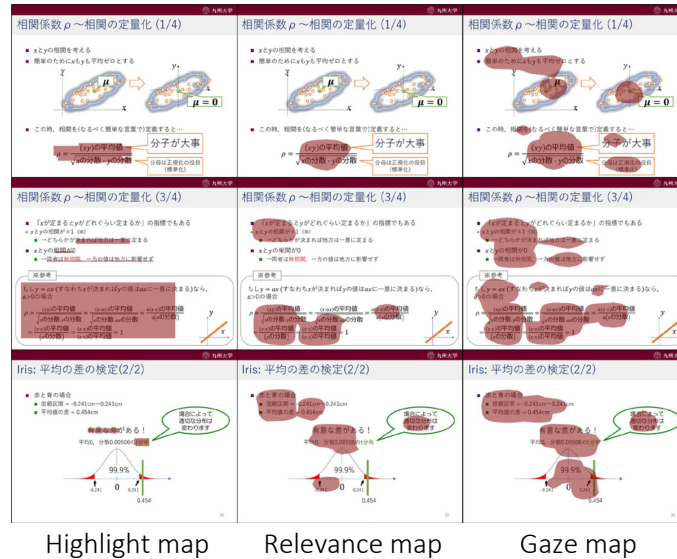


Figure 1: Highlight maps, relevance maps, and gaze maps for the three most difficult pages

Table 1: Averaged precision (P), recall (R), and F-measure (F) values

	Highlight vs. relevance			Highlight vs. gaze		
	P	R	F	P	R	F
Top 3 difficult pages	0.483	0.497	0.394	0.292	0.749	0.334
The other pages	0.108	0.316	0.129	0.098	0.604	0.150

that is related to page difficulty and students' attention. Therefore, we assert that highlights and eye movements can be complementarily combined for estimating page difficulty. For instance, when a page area is detected in both a highlight and a relevance map, this suggests that the area has higher confidence than areas that were detected in either a highlight map or a relevance map. In addition, the system can analyze lower confidence areas based on the features of higher confidence areas, thus rejecting or accepting the former. We have concluded that a combination of learning logs from different resources can be useful for such an analysis.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Number JP19K20421 and the MEXT Society 5.0 Realization Research Support Project Grand Challenge Research in Osaka University, Japan.

REFERENCES

- Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K. R., & Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS one*, 10(7), e0130140.
- Minematsu, T. (2019). Region-wise page difficulty analysis using eye movements. *Proceedings of 16th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2019)*.
- Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-Book-based learning analytics in university education. *Proceedings of the 23rd International Conference on Computer in Education (ICCE 2015)*, pp. 401-406.

Reduction of Supervised Data for Chat Analysis in Collaborative Learning by Using Transfer Learning Methods

Shun Moriya, Chihiro Shibata, Kimihiko Ando, Taketoshi Inaba

Tokyo University of Technology

c0115334ef@edu.teu.ac.jp, {shibatachh, ando, inaba}@stf.teu.ac.jp

ABSTRACT: With increasing interest in Learning Analytics for big educational data in recent years, studies on utilization of deep learning, have been performed actively. Our research team has also constructed AI models that automatically and in real time add coding label to collaborative learning data, verified its accuracy and its usefulness. However, enormous manpower required for preparing supervised data has been focused on as a problem. And, it may further increase cost to give new coding labels that correspond to individual needs of educational practices. Therefore, in this study, we pursue how amount of human labeling workload can be reduced by utilizing transfer learning. In the proposed method, a language model is pre-trained in advance to reduce the amount of the supervised data.

Keywords: CSCL, coding scheme, automatic coding, deep learning, transfer learning

1 BACKGROUND AND PROBLEM PRESENTATION

Our research group has built up a computer-supported collaborative learning (CSCL) system on a Moodle which works as LMS in the university and analyzed educational data obtained in actual lectures. In our previous study, 16 coding labels to simply indicate speech act characteristics of each statement have been developed aiming at analyzing collaborative learning process for part of data among more than 30,000 statements of data collected during collaborative learning and performed labeling (coding) by manpower. In addition, it has been demonstrated in the previous study that coding label is predicted with an accuracy exceeding 0.7 by κ coefficient by making learn labeling using deep learning based on such data (Ando, Shibata & Inaba, 2017).

It is possible to perform automatic coding with high accuracy using deep learning technology for chat data as long as sufficient amount of statement data with coding labels is prepared. However, it is required to prepare enormous statements and coding labels in order to utilize deep learning technology. It is also required to be conducted by well-trained coders for obtaining results as just intended, resulting in necessity of enormous human resources. In order to resolve these issues, through the experiments, we verify that amount of coding labels required to be given by actual human work can significantly reduce by the proposed method. We also show that sufficient accuracy can be maintained even based on less data.

2 CODING SCHEME AND DATASET

In reference to a multidimensional coding scheme proposed by Weinberger et al., a scheme consisting of four dimensions has been adopted in the study (Weinberger, & Fischer, 2006). Coding label was given by unit of statement in chat. A code of each of four dimensions is selected from among plural

labels and given. As detail of each dimension has been already described in detail in the previous study, just contents of each dimension and labels are shown in Table 1 (Shibata, Ando & Inaba, 2018). In addition, chat data for learning here to be analyzed in the study was obtained from our own CSCL system. Number of group members and students, total number of groups and statements as well as time of discussion in the chat data extracted from seven lectures are 3 ~ 4, 426, 202, 9,962 and 45 ~ 90 minutes, respectively. The supervised data was obtained by giving four dimensions label for each statement by manual work by two coders.

Table 1: Overview of Coding Scheme.

Dimension	Description	Labels
Epistemic	How to be directly involved in problem solving	On task, Off task
Argumentation	How to construct arguments and counterarguments	Simple claim, Qualified claim, Grounded claim, Qualified and Grounded Claim
Social	How to cope with others' statements	Externalisation, Elicitation, Quick consensus, Integration-oriented consensus
		Conflict-oriented consensus, Summary
Coordination	How to coordinate to advance discussion smoothly	Task division, Time management, Technical coordination, Proceedings

3 PROPOSED METHOD

We show the flow of the transfer learning in Figure 1. First, a language model is learned using a huge data set A without labels (Pre-training). Thereafter, a data set B having a small scale but with coding labels is prepared. Then, using B, the deep neural network is learned so as to classify the coding labels (Supervised learning). In the supervised learning, the weights of the model already learned in A is used as a part of the model for B. In this study, we used a total of approximately 30,000 statements, of which approximately 10,000 were labeled and the remaining 20,000 were unlabeled. Furthermore, intentionally we reduce the number of B, and verify through the experiments whether the coding accuracy can be maintained to a certain degree with a small amount.

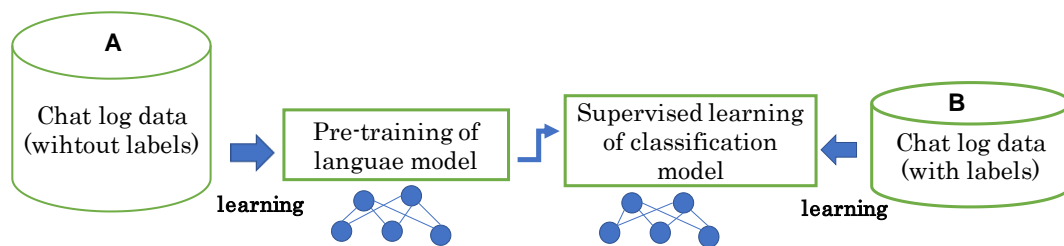


Figure 1: Flow of the Transfer Learning

The deep learning model used in the study consists of a combination of multi-stage recurrent neural networks (RNN). An architecture called long short-term memory (LSTM) is used for RNN, which is well known for its excellent predicting capability in learning time-series data including natural languages. In the pre-training, a part of the model for supervised learning is learned in advance. At first, learning is performed by inputting sentences in forward direction. In the supervised learning, instead of using all the output vectors, each of which is corresponding to each word in the sentence, we use only the output vector corresponding to the final word to classify the coding labels. On the other hand, in the pre-training, for each word, the output vector of the RNN block is used to calculate the probability of the next word. The pre-training in the reverse direction is performed in a similar way to the training in the forward direction.

4 EXPERIMENT AND EVALUATION

24,407 statements with vocabulary of 9,219 words for pre-training of language model are also extracted from chat data of lectures different from those of the supervised learning. However, labels have not been given to these statements. While number of statements in the supervised learning is different by the dimension, number of statements of each Epistemic, Argumentation, Coordination, and Social dimension is up to 7,614, 6,988, 3,159 and 2,357, respectively. These numbers are different because how often these labels are given varies from dimension to dimension. Experiments are performed by changing the number of the coding labels, or the amount of supervised data. We assume 20--25 data sizes for each coding label for the experiments. Prediction accuracies are obtained based on 10-fold cross validation for each coding dimension and data size. 940 sets of experiments were performed in total. Comparison of error rate is performed between cases with our proposed pre-training (transfer) method and without pre-training (scratch). As shown in Figure 2, using pre-learning, the error rates reduced in the three dimensions except for Coordination dimension. In these three dimensions, difference between cases with and without pre-training in error rate tends to increase as the size of supervised data decreases.

5 CONCLUSIONS

It has been suggested that our pre-training may quite significantly reduce amount of supervised data. It has been also revealed that the effect may differ depending on dimensions. These results indicate it possible to drastically reduce human resources cost for preparing supervised data which is a big hurdle for further introduction of automatic coding by way of our proposed method. Furthermore, it can be said that the proposed method could make LA using the coding methods easier than before. It is the next step for us to perform accuracy verification based on the approach here adopting other multi-dimension coding schemes with structural complexity at similar level to that in the current study.

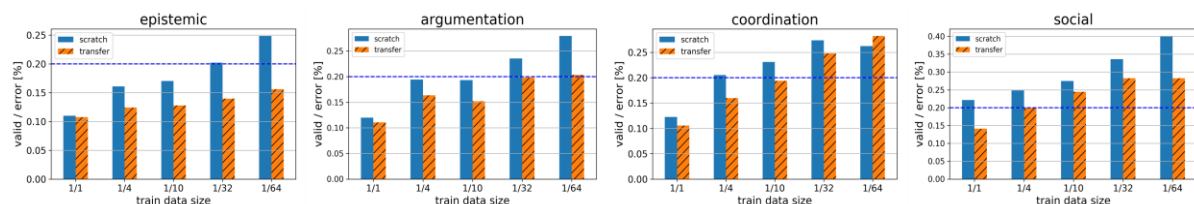


Figure 2: Effect of proposed method by data size. Vertical axis indicates error rate (1 – Correct answer rate). Left bars: without proposed method(scratch), Right bars: proposed method(transfer).

REFERENCES

- Kimihiko Ando, Chihiro Shibata, Taketoshi Inaba (2017). Coding Collaboration Process Automatically: Coding Methods Using Deep Learning Technology. *International Journal on Advances in Intelligent Systems*. 10(3&4), 345-354.
- Weinberger, A., & F. Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported learning. *Computer & Education*, 46(1), 71-95.
- Shibata Chihiro, Ando Kimihiko, Inaba Taketoshi (2018). *Evaluation of Automatic Collaborative Learning Process Coding Using Deep Learning Methods*. In proceedings of IEEE TALE 2018 : International Conference on Engineering, Technology and Education, 2018 , 251-258.

Precision-Based Predictive Analytics

Patsy Moskal

University of Central Florida
Patsy.Moskal@ucf.edu

Thomas Cavanagh

University of Central Florida
cavanagh@ucf.edu

Morgan Wang

University of Central Florida
Chung-Ching.Wang@ucf.edu

Jianbin Zhu

University of Central Florida
zhujianbin@gmail.com

ABSTRACT: Early, accurate identification of at-risk students maximizes the opportunity for effective intervention. We examined students' progression through key courses at a large, metropolitan U.S. university with a significant online course presence, predicting student risk using variables from both the student information system (SIS) and course learning management system (LMS). Initially, hierarchical logistic regression models were developed to serve as the baseline using student information system (SIS) data. Then, weekly progression models were developed using both SIS and LMS data. Training was based on student data captured as they progressed through courses during Fall 2017 with test data from students enrolled in Fall 2018. The results of at-risk students for each week were identified and compared. Model evaluation showed the baseline models can predict at-risk students with a high level of precision in week one. The baseline models' performance was confirmed using weekly iteration. In addition, weekly progression models provide insight about factors leading to the reduction of students' academic risk. Because this model provides early identification, findings can be beneficial for faculty and others for developing intervention support and resources to help at-risk students ultimately succeed.

Keywords: predictive modeling, data mining, at-risk students, learning management system, student information system, learning analytics, academic performance

1 INTRODUCTION

Research to identify academically at-risk students often uses a variety of methods (Smith et al., 2012; Wladis et al., 2014; Miguéis et al., 2018; Simanca et al., 2019). This study makes a case for the use of *model precision* as the best selection criterion, instead of the more commonly used accuracy model. For example, assume that one class has 250 students and a 30% failure rate. The confusion matrices for Decision A and Decision B are seen in Table 1(a) and Table 1(b), respectively. The accuracy rate for Decision A, 80%, is higher than the accuracy rate for Decision B, 78%. However, Decision B has a 60% precision rate and allows us to identify 20 more academically at-risk students than Decision A because of its consistency.

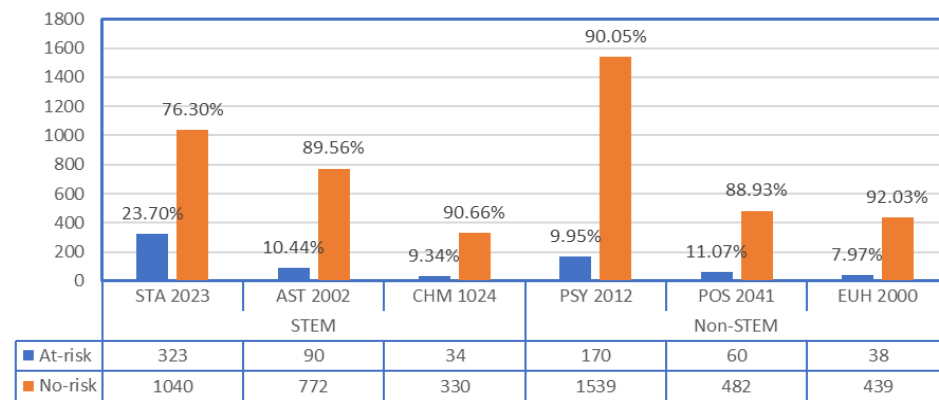
Table 1: Examples of two different precision rates for decisions

(a) Confusion Matrix for Decision A				(b) Confusion Matrix for Decision B			
True State	Predict State		Row Total	True State	Predict State		Row Total
	Pass	Fail			Pass	Fail	
Pass	175	0	175	Pass	150	25	175
Fail	50	25	75	Fail	30	45	75
Column Total	225	25	250	Column Total	180	70	250

In addition, most prior studies use training and validation data collected during the same semester with compressed time lag. Note that the predictive model cannot be built at the beginning of any semester since the target variable (i.e., the student's final course grade) is not known at that time. A useful model should be built in one semester and then validated using data collected in future semesters. Otherwise, the performance measure presented is not "honest." Typically, this measure could be biased, and overestimate precision.

2 BACKGROUND

The purpose of this study was to identify academically at-risk students using analytics early enough in the semester to successfully intervene and reduce students' academic risk. We used data mining to perform predictive analysis, incorporating student and course data from both the student information system (SIS) and the learning management system (LMS). Course data collected in Fall 2017 was used to develop the model for courses and course data collected in Fall 2018 was used to validate the developed models. The SIS system provides students' enrollment information, including demographics, academic type and level, GPA and academic test history (SAT, ACT). LMS data for each course included course sections, student engagement activities and performance records. Courses from both STEM and non-STEM were considered, including those with significant online components and that were part of the general education program (GEP). The distribution of the number of at-risk students for six selected courses is shown in Figure 1.

**Figure 1: Students at risk by course.**

3 PREDICTIVE MODELING

In our study, one baseline hierarchical logistic regression model was developed at the beginning of each semester and one weekly progression model was developed each week from week 2 to week 6 for STEM and non-STEM courses separately. To ensure these models were generalizable, they were built using course and student data from fall 2017 and then validated/tested using data from fall 2018 course iterations. For the baseline model, only SIS data were used. Both LMS and SIS data were used to build the weekly progression models. Three STEM courses (Statistics STA 2023, Physics AST 2020,

and Chemistry CHM 1024) and three Non-STEM courses (Psychology PSY 2012, Political Science POS 2041, and Western Civilization EUH 2000) were selected for this study as they were part of the general education program and had corresponding online sections that might more heavily use the LMS.

For the baseline model, the target variable was students' academic risk (binary variable based on course grade). Predictors considered included both demographic and academic variables. Demographic variables included gender, ethnicity, academic level (freshman, sophomore, junior, senior), full/part time enrollment indicator, first generation college student indicator, transfer student indicator, and birthdate. Academic variables included high school GPA, institutional GPAs (current semester and cumulative), SAT score, and ACT score. Only a subset of these predictors was selected by the baseline procedure. In addition, all information from the LMS gradebook was organized into three types of scores: assignment, quiz and exam, and then combined with SIS variables to build weekly progression models for week 1 to week 6. Six predictive models were built for STEM courses and six weekly progression models were built for non-STEM courses.

4 RESULTS

This research produced some noteworthy findings, some presented in Table 2: (1) the precision rate for the baseline model using SIS data alone is approximately 85%, i.e., 85% of at-risk students identified by the baseline model failed the course at the end of the semester; (2) the precision rate of the baseline model is similar to the weekly progression model week 2 to week 6; (3) the LMS data reveal that the top three factors for students at risk at the beginning of the semester who eventually pass the course are GPA, Exam, and Quiz scores for STEM courses; (4) the baseline model works for both STEM courses and non-STEM courses; however, the precision rate for STEM courses is 5% to 10% higher; (5) the factors leading to students eventually passing the course can be used to develop preventative interventions that may reduce students' academic risk. Future research will investigate the use of these findings to inform course design through faculty development initiatives and examine possible student intervention methods to improve student success.

Table 2: Examples of Precision Rates of Students' At-risk Designation from Baseline and Weekly Progression Predictive Models

Course	True	Baseline	Week 2	Week 3	Week 4	Week 5	Week 6
STEM	At-risk	62	62	62	62	63	64
STA2023	No-risk	8	8	8	8	7	6
(Cut 70 At-risk)	Precision	88.57%	88.57%	88.57%	88.57%	90%	91.4%
Non-STEM:	At-risk	40	41	41	41	41	46
PSY2012	No-risk	10	9	9	9	9	4
(Cut 50 at-risk)	Precision	80%	82%	82%	82%	82%	92%

REFERENCES

- Smith, V. C., Lange, A., & Huston, D. R. (2012). Predictive modeling to forecast student outcomes and drive effective interventions in online community college courses. *Journal of Asynchronous Learning Networks*, 16(3), 51-61.
- Wladis, C., Hachey, A. C., & Conway, K. (2014). An investigation of course-level factors as predictors of online STEM course outcomes. *Computers & Education*, 77, 145-150.
- Miguéis, V. L., Freitas, A., Garcia, P. J., & Silva, A. (2018). Early segmentation of students according to their academic performance: A predictive modelling approach. *Decision Support Systems*, 115, 36-51.
- Simanca, F., González Crespo, R., Rodríguez-Baena, L., & Burgos, D. (2019). Identifying Students at Risk of Failing a Subject by Using Learning Analytics for Subsequent Customised Tutoring. *Applied Sciences*, 9(3), 448.

Toward Predicting Learners' Efficiency for Adaptive e-Learning

Yuta Nakashima¹, Hirokazu Kobori², Ryota Takaoka¹, Noriko Takemura¹, Tsukasa Kimura¹,
Hajime Nagahara¹, Masayuki Numao¹ and Kazumitsu Shinohara¹

¹Osaka University

{n-yuta, takaoka, takemura, nagahara}@ids.osaka-u.ac.jp

{kimura, numao}@ai.sanken.osaka-u.ac.jp, shinohara@hus.osaka-u.ac.jp

²DAIKIN INDUSTRIES, LTD.

hirokazu.kobori@daikin.co.jp

ABSTRACT: e-Learning can be a feasible and promising approach to providing more learning opportunities. One main criticism of e-learning may be its absence of teachers/tutors, who can infer the learner's comprehension and add a more detailed explanation on a topic the learner may not fully understand. In this work, we aim at *adaptive e-learning*, which automatically adjust the content according to the learner's comprehension level to each topic, and introduce initial results on predicting learners' efficiency using an eye tracker.

Keywords: e-learning, eye tracker, deep neural networks, learner's efficiency

1 INTRODUCTION

Estimating how much a learner is engaged, concentrates, and comprehends the topic, which we call *learners' efficiency*, is an essential step towards adaptive e-learning sessions to automatically adjust the content or suggest resting. In this paper, we explore the first step toward estimating learners' efficiency with a certain learning task and a definition of learners' efficiency. One essential question is what signals from the learner we can use. From various possible signals, e.g., vitals, EEG signals, facial expression, we choose eye trajectories captured by an eye tracker as eye trajectories are one of the modalities that can be directly affected by the learner's status. We formulate the learners' efficiency estimation as a binary classification problem and solve it with a deep neural network.

2 TASK DESIGN

For testing eye tracking-based learners' efficiency estimation, we design a small read-and-recall task, in which a learner (i) reads in their own pace a short text (1,600-1,800 words in Japanese) on a certain topic divided into 8 pages and displayed on a screen, and (ii) answers as quick as possible to 12 questions asking if a certain word appears in the text. Six of them appear in the text and the other six do not (i.e., new words). Among the words that appear in the text, three are general words (e.g., "we"), and the others are specific to the topic of the text. Both types of words are extracted based on each word's frequency and co-occurrence rates: General words have a low frequency and co-occurrence, whereas those of specific words are high. Correctly answering to the general word questions is tough compared to specific word questions. In a single session, a learner works on 15 trials of the task along with other tasks involving short-term memory and questionnaire. The learner participates in another session after 30-min. rest. A single session roughly takes 60 min. A single learner is asked to complete these two sessions in a day and to be engaged for two days, which are not necessarily consecutive,

where each session is done in different environmental conditions (various temperature and humidity). The details of the task can be found in [Kimura, 2019].

3 DEFINITION OF LEARNERS' EFFICIENCY

Upon this task, we define two criteria for learners' efficiency over a single trial. The first criterion (a) is directly derived from the performance of each trial. Formally, let E_d equal to N_c/N where $N = 12$ is the number of questions and N_c is the number of correctly answered questions. We can define this ratio as the learner's efficiency and regress it; however, due to the difficulty in regression, we formulate the learners' efficiency estimation as a binary classification problem. We assign to each trial a label *efficient* when $E_d \geq 10/12$ and *non-efficient* when $E_d \leq 8/12$. The trials whose $E_d = 9/12$ are discarded to balance the label distribution as well as reduce uncertainty.

The second criterion (b) considers the times when reading the text and answering the questions as well. Let v be a vector for each trial that contains (i) the working memory test result (a dot is displayed in a cell of a five by five grid consecutively and the learner clicks in the cells that showed the dot, where the correct answer rate is used as the result); (ii) the time to read the text, the correct answer rates for (iii) general, (iv) specific, and (v) new words; and the average times until answering for (iv) general, (vii) specific, and (vii) new words; (ix) overall correct answer rate; and (x) the overall time to answer. We applied Ward's method [Ward, 1963] to cluster v 's into groups of higher performances and faster reading/recall times and of lower performances. We label the trials in the first group as *efficient* and the other group's as *non-efficient*.

4 PREDICTING LEARNERS' EFFICIENCY WITH EYE TRACKING

For predicting learners' efficiency, we use eye trajectories on the screen. Let $p \in \mathbb{R}^{2 \times K}$ be a concatenation of K positions $(x_k, y_k)^T$ on the screen. The number of positions K can differ from trial to trial. Our binary classification task is to find a function that maps p to label *efficient* or *non-efficient*. We use deep neural networks. Specifically, we use a convolutional neural network with either $L = 1, 2$, or 3 convolution layers, each of which is followed by ReLU nonlinearity. The output from the top convolution layer is then mean- or max-pooled to get a single feature vector and is fed to a fully-connected layer with softmax nonlinearity. Figure 1 shows our network architecture when $L = 2$. For training, we use dropout before the fully-connected layer and Adam [Kingma & Ba, 2015]. Since eye trackers may fail to track the eyes, we either fill by zeros or interpolate missing values with the second-order spline interpolation.

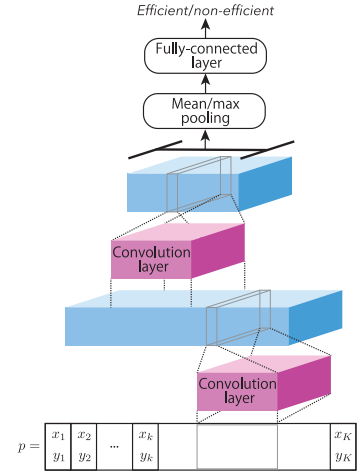


Figure 1: Network architecture when $L = 2$.

5 RESULTS

For evaluating our model, we recruited 21 participants mostly in their 20s, who were asked to have two sessions. We used Tobii Pro X3 120 as our eye tracker to obtain eye trajectories. The eye tracker tracks left and right eyes simultaneously, and we took the average over the two. Each convolution layer's kernel size W is either 60, 120, or 240. Since the number of trials is small, we adopted leave-

one-out cross-validation: The model is trained with 20 participants' trials and evaluated with the other 1 participant, which is repeated until all participants' are used for evaluation. The numbers of trials with efficient and non-efficient labels are 880 and 515 for the first criterion (a), and are 1,189 and 205 for the second criterion (b). Figures 2(top) and (bottom) show the box plots of 21 AUCs for (a) and (b), respectively, in different configurations. No significant differences are observed for (a), which almost failed in predicting the efficiency. Meanwhile, Figure 2(bottom) shows the average AUCs of around 70% when $L = 2$ and 3 with mean pooling. Again, there is no significant difference in performance for different kernel sizes. Although the reading/recall time is correlated to the labels under criterion (b), we believe that our model can predict the learners' efficiency from eye trajectories because convolution layers are location invariant.

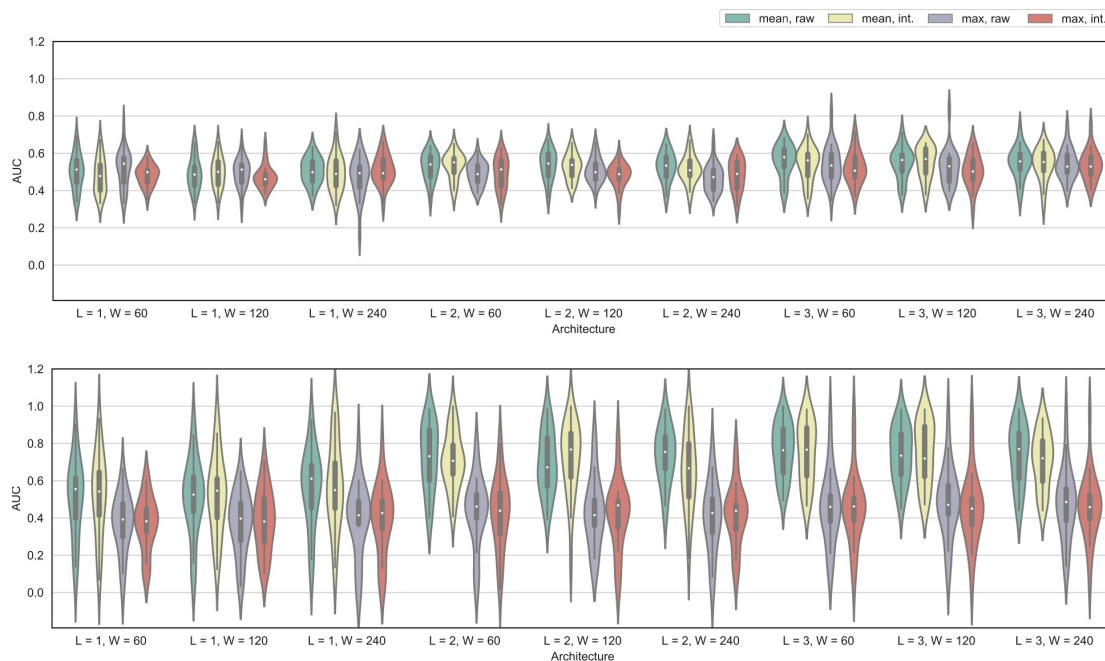


Figure 2: Violin plots of the AUCs for the first (top) and second (bottom) criteria.

6 CONCLUSION

This paper presents our first attempt toward learners' efficiency estimation using eye trajectories and a deep neural network. The experimental results are encouraging. We will refine the task design and investigate why our second criterion with mean pooling worked. Besides, our current model works only after one trial is done, but we will explore to make it real-time. This work was supported by Daikin Industries, Ltd.

REFERENCES

- Kimura, T. et al. (2019). Heat environment increases implicit mental workload even if learning efficiency is enhanced. Proc. The 3rd Annual Convention of the Japanese Psychological Association. (in Japanese)
- Kingma, D.P., Ba, J. (2015). Adam: A method for stochastic optimization. Proc. International Conference on Learning Representations (ICLR). <https://arxiv.org/abs/1412.6980>
- Ward, J. H., Jr. (1963), Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association, 58, 236–244.

Simultaneously Learning Competencies and Item Difficulties

Kai Neubauer and Ulf Brefeld

Leuphana University

{kai.neubauer,brefeld}@leuphana.de

ABSTRACT: Calibrating itempools for computerized adaptive testing is often tedious and expensive. As a remedy, we propose to learn abilities and item difficulties simultaneously. Since the estimation of person and item parameters is scale-free, the goal is to compute a ranking of examinees, rather than a precise estimation of competency. We report on results from a simulation study for Rasch models and 3PML, respectively.

Keywords: CAT, item difficulties, Rasch model, 3PLM

1 INTRODUCTION

The proficiency of a person is a latent trait. It cannot be observed directly and has to be measured by appropriate means like a test. Computerized adaptive tests (CATs) aim to estimate a persons' proficiency by optimizing the item selection. Compared to traditional linear tests, CATs can increase the accuracy of the proficiency estimates with shorter test lengths (Wainer, Dorans, Flaugher, Green & Mislevy 2000). However, these advantages come at a cost: An essential requirement for a CAT is a calibrated itempool that often results from a tedious and expensive calibration study. We study two CAT approaches that estimate proficiencies without the need for calibrated itempools by trading the assessment of exact abilities for a ranking of participants. The first approach relies on a *constrained joint maximum likelihood* estimation of the Rasch model (Rasch, 1960) while the second leverages the *Bradley-Terry model* (Bradley & Terry, 1952) to compute a ranking of the examinees based on pairwise comparisons between their responses to the same items.

2 RANKING EXAMINEES

2.1 Problem Setting

Let Y_{ij} be the event of examinee i responding to item j with outcome $y_{ij} \in \{0,1\}$ where $y_{ij} = 1$ indicates a correct and $y_{ij} = 0$ an incorrect response. We assume that the response behavior of examinee i is completely determined by their ability $\theta_i \in \mathbb{R}$ and that responses Y_{i1}, \dots, Y_{iJ} of the same examinee are conditionally independent. Furthermore, items are presented to examinees in synchronized rounds that start and end simultaneously for all candidates. Variable $\omega_{ij} \in \{0,1\}$ indicates whether item j has been presented yet to the i -th examinee ($\omega_{ij} = 1$) or not ($\omega_{ij} = 0$). In each round, a selection algorithm determines individual items that will be presented to the examinees. The algorithm has access to all previous responses up to that round. The final objective is to create a ranking of the examinees regarding their ability such that higher ranks correspond to higher abilities.

2.2 Ranking via Ability Parameter Recovery

The first approach is similar to the idea of item calibration. We use a constrained joint maximum likelihood (CJML) estimator (Chen, Li & Zhang, 2019) that relies on the Rasch model. The applied constraints limit the parameter spaces of difficulties and abilities and render parameter estimation feasible even for constantly correct ($y_{ij} = 1$) or incorrect ($y_{ij} = 0$) responses where $\omega_{ij} = 1$ for all examinees i or items j , respectively. The objective function is given by

$$\arg \min_{\theta_i, b_j} \sum_{i,j: \omega_{ij}=1} \log(1 + \exp(\theta_i - b_j)) - y_{ij}(\theta_i - b_j) \quad s. t. |\theta_i| \leq C, |b_j| \leq C.$$

Since the sets of abilities and item difficulties are separable, the objective is *optimized via alternating minimization* where intermediate solutions are projected to satisfy the simplex constraints on θ and b , respectively. The item selection is based on a two-phase design (Makransky & Glas, 2014). In the first phase, we select items in a controlled random design to increase the number of responses for each item uniformly up to a predefined threshold. In the second phase, items are selected with respect to the maximum item information. Therefore, item and person parameters are re-estimated after every round. The final ranking is computed by sorting the estimated abilities.

2.3 Ranking via Pairwise Comparisons

The Bradley Terry Model (BTM) expresses the probability $P(i > j)$ that item i is preferred over item j by assigning a strength parameter $s \in \mathbb{R}^+$ to each item. The strength parameters are estimated using pairwise comparisons between the items and can be used to derive a ranking. Assuming that pairwise comparisons between items are independent, the strength parameters for a set of N items can be estimated by *majorize-maximization* (Hunter, 2004)

$$\arg \max_{s_i} \sum_{i=1}^N \sum_{j=1}^N (a_{ij} \ln(s_i) - a_{ij} \ln(s_i + s_j)),$$

where a_{ij} denotes the number of times item i is preferred over object j . In our case, we identify responses to the same item with a pairwise comparison of the two examinees. Since the test only consists of dichotomous items, a pairwise comparison either expresses a distinct preference for one examinee, or there is a tie if both examinees respond identically. To account for ties, a_{ij} is replaced by a scoring function. We expect the accuracy of the ranking to improve as the number of comparisons increases. However, not every pairwise comparison is equally useful for determining a ranking since ties do not provide any information about the order of examinees. Therefore, our selection algorithm prefers items with a balanced ratio of correct and incorrect responses since this maximizes the expressivity of the approach from a population-wise perspective.

3 SIMULATION STUDY

The performance of the proposed approaches is evaluated with a simulation study in which responses are generated according to Rasch and 3PL models, respectively. We assume an uncalibrated itempoolⁱ for the proposed methods and include a conventional CAT with calibrated itempoolⁱⁱ as baseline. We simulate a CAT with 300 examinees and 200 items and report on averages over 400 repetitions. Figure 1 shows the proportion of the correctly selected top 20% of the participants over the course of the test. The dashed line indicates the transition between the item selection phases and the shaded areas represent the 95% confidence interval.

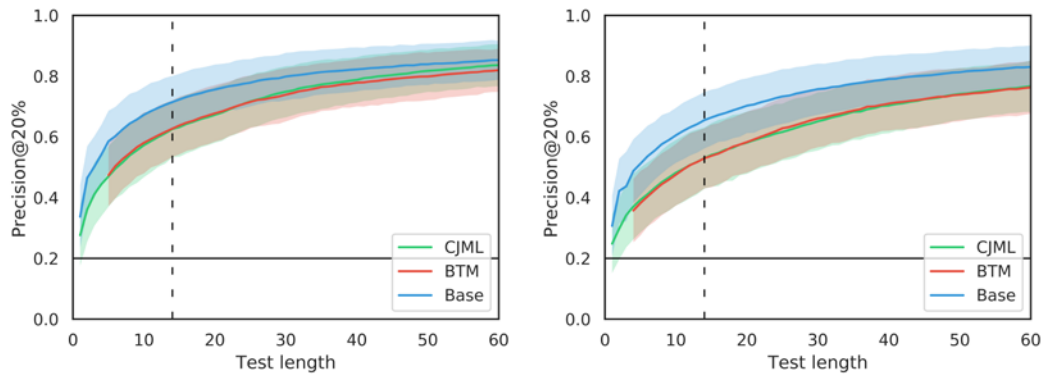


Figure 1: Results with an uncalibrated itempool for the Rasch model (left) and 3PLM (right)

As test lengths increase, rankings improve across all methods but saturate before a fully correct solution can be recovered. We attribute this finding to the probabilistic response behavior that render distinguishing similarly performing individuals difficult. While longer tests lead to smaller differences between the proposed methods, the baseline significantly outperforms the BTM and CJML. However, recall that the baseline requires a fully calibrated item pool and thus leverages much more information while our approaches start from scratch. Additional experiments reveal that the performances of our approaches depend on the characteristics of the itempool. For example, we observe that the performance is inverse proportional to the number of items with random response patterns. These items render the prediction problem difficult. The baseline clearly benefits from identifying these items already in the calibration study. However, the experiments also show that both proposed methods effectively identify and exploit useful items and lead to appropriate rankings.

REFERENCES

- Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, 39(3/4), 324-345. <https://doi.org/10.2307/2332730>
- Chen, Y., Li, X., & Zhang, S. (2019). Joint maximum likelihood estimation for high-dimensional exploratory item factor analysis. *Psychometrika*, 84(1), 124-146. <https://doi.org/10.1007/s11336-018-9646-5>
- Hunter, D. R. (2004). MM algorithms for generalized Bradley-Terry models. *The annals of statistics*, 32(1), 384-406. <https://doi.org/10.1214/aos/1079120141>
- Makransky, G., & Glas, C. A. (2014). An automatic online calibration design in adaptive testing. *Journal of Applied Testing Technology*, 11(1), 1-20.
- Rasch, G. (1960). Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests.
- Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., & Mislevy, R. J. (2000). *Computerized adaptive testing: A primer*. Routledge. <https://doi.org/10.4324/9781410605931>

ⁱ We assume for the item response model parameters: discrimination $a \sim \text{Lognormal}(0, 0.5)$, difficulty $b \sim N(0, 2)$, pseudo-guessing $c \sim N(0.25, 0.02)$. Further, we assume that 10% of all items are corrupted i.e. the probability for a correct answer does not depend on the ability, but chance. The probability for a correct response is given by a $\text{Beta}(1, 15)$ distribution.

ⁱⁱ Each item is calibrated with 500 responses prior to the test. The itempool does not include corrupted items and the item discrimination parameters follow a $N(1.2, 0.25)$ distribution. The baseline uses the Bayes modal estimator for scoring abilities and maximum item information for selecting items.

A Qualitative Analysis of One University's Ethical Fears and Practical Desires for Learning Analytics

Melanie E. Pepper

University of Colorado
melanie.pepper@colorado.edu

Jessie Sutton

University of Northern Colorado
sutt5261@bears.unco.edu

ABSTRACT: As higher education embraces technology and learning analytics the amount of available student data grows exponentially. Although the potential for new insights about student learning is exciting, there has been very little work done on what students, one of the major stakeholders and potential benefactors of learning analytics data, know about learning analytics. Here we sought to assess the state of knowledge around learning analytics among the end users, namely university students and faculty. We conducted focus groups and performed a grounded theory analysis to identify three emergent themes about what end users at one mid-sized research institution know about learning analytics and what types of best practices students and faculty would like to see implemented. As the field of learning analytics rapidly expands, the insights generated by this work are important for tempering new research insights with what the end users want and need from the technology.

Keywords: Learning analytics, policy, education, professional development

1 INTRODUCTION

The use of technology in higher education continues to expand, and with each new technology, there are new opportunities for learning analytics. The potential for learning analytics to revolutionize education is undisputed. However, there is a lack of specific data training for many educators (Tsai et al., 2018) and although learning analytics are intended to create a better environment for students, very little research has been done that include students' perspectives (Gasevic, Dawson, & Siemens, 2015; Prinsloo & Slade, 2017). Tsai and colleagues (2018) also point out that all stakeholders must be engaged in and buy into learning analytics for it to be effective. Given the importance of end user viewpoints for receiving full utility of benefits from learning analytics and the lack of knowledge about what end users know and think about learning analytics, we conducted focus groups to determine what students and faculty know about learning analytics. We use these findings to make recommendations for student and faculty-informed best practices for learning analytics in higher education.

2 METHODS

Our study included 29 participants from a single mid-size public doctoral granting institution located in the Rocky Mountain Region of the United States of America. Participants represented several university departments (e.g. biology, speech and language sciences, statistics) and included 19

students and 10 faculty/staff. Approximately four individuals at a time participated in a focus group, during which a member of the research team led participants through a questioning route. The questioning route was designed to probe what individuals already knew about learning analytics and big data as well as broader ethical questions about data ownership and informed consent. Participant responses were recorded using a digital audio recorder and then professionally transcribed. To identify emergent themes, we utilized a qualitative, grounded theory approach (Creswell, 2014) with QSR-NVivo (Hutchinson, Johnston, & Breckon, 2010). As is customary with a grounded theory approach, both authors first independently reviewed all transcripts and identified key themes. After initial review, the authors came back together to compare observations and mutually decide upon final themes. We addressed qualitative reliability via modeling our analysis off of previous work (Creswell, 2014; Hutchinson, Johnston, & Breckon, 2010). As a check of qualitative validity, a peer de-briefer reviewed our analysis and an external auditor who is not a member of the research team or familiar with the project reviewed the entire project (Creswell, 2014).

3 RESULTS AND DISCUSSION

3.1 Theme 1: General Lack of Learning Analytics Knowledge

First, we noted that members of the university community had an incomplete understanding of learning analytics. For example, there was an awareness of rudimentary analytics, but how to interpret those analytics, either as a student or professor, was lacking. One participant mentioned that they used time spent logged into the learning management system as a method for determining the amount of assistance they would offer a student, assuming that a student who spent more time logged in was also more engaged. This is not necessarily a valid benchmark and misinterpretations of basic analytics could lead to classroom inequity. Although time spent logged in is low-level analytics, as the field continues to grow and more complicated analytics become prevalent, it is important that both students and faculty understand how to interpret and use this information. Furthermore, as learning analytic software and methodologies expand, unfamiliarity with learning analytics can spark fear and distrust, which creates potential roadblocks for effective implementation (Drachsler & Greller, 2016). Therefore, it is important that there are educational opportunities for end users to understand what learning analytics are and how they can be used effectively.

3.2 Theme 2: Necessity of Learning Analytics Education and Professional Development for Students and Faculty

As noted earlier, faculty had some familiarity with analytics available to them via the learning management system. However, the lack of training and understanding of how to use big data in the classroom can lead to potentially worrying concerns about equity. Learning analytics data can yield new perspectives on students, but it is important for educators to learn that it does not give us every perspective and intellectual efforts will not always translate to digital metrics. Training instructors on how to interpret data may help promote equitable outcomes and avoid excluding students based on misinterpretation of data. In addition, we encourage faculty to examine data sharing policies for a given technology prior to classroom adoption. It may be worthwhile for university information technology offices to provide some guidance to faculty on best technologies to use, or things to be aware of when choosing a technology for the classroom.

We noted that students were generally unaware of how much data they generate and what data is collected. Participants felt consent was important (discussed below) but many were fairly flippant when it comes to granting entities access to use their data. Furthermore, as noted by Rubel and Jones (2016), the technology and algorithms involved in learning analytics are so complex that informed consent feels impossible for those who do not understand them, which is the majority of research participants at this institution. As such, an important first step is creating opportunities for students to learn about big data. One participant stated, *"We need to start teaching that...information is commodified now...I think because it's so new that we haven't quite started that, it's going to happen and we just need to become better, financially responsible students or academics as early as possible."* One participant suggested including big data education during new student orientation. Possible future directions could examine the efficacy of methods to teach end users about what learning analytics are, proper usage, and translation into actionable insights.

3.3 Data Awareness, Usage, and Informed Consent

We observed an overall lack of consensus among our participants regarding how to manage the issue of consent around learning analytics data. Some felt that matriculating at the University was akin to giving your consent for data to be used while others felt an informed consent process was necessary. Participants agreed that an important first step to addressing issues of data awareness and informed consent was for universities to be transparent with students about what data is being collected, how that data is used by the university and/or third parties, and student rights about their data, such as the right to be forgotten. Participants felt that transparency was important for fostering trust and will facilitate the implementation of learning analytics initiatives, which is consistent with recommendations in Drachsler and Greller (2016).

REFERENCES

- Creswell, J.W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Thousand Oaks, CA: SAGE Publications.
- Drachsler, H., & Greller, W. (2016). Privacy and learning analytics – it's a DELICATE issue. In *LAK'16*, April 20-24, Edinburgh, UK, ACM 978-1-4503-3417-4/15/03. <http://doi.org/10.1145/27233566.2723642>
- Gasevic, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. In *Tech Trends* 59(1), pp 64-71.
- Hutchinson, A., Johnston, L., & Breckon, J.D. (2010). Using QSR-NVivo to facilitate the development of a grounded theory project: an account of a worked example. *International Journal of Social Research Methodology*, 13:4, 283-302. DOI: 10.1080/13645570902996301
- Prinsloo, P., & Slade, S. (2017). Ethics and learning analytics: charting the (un)charted. *Handbook of Learning Analytics*, Edition: 1st, Chapter: 4, Publisher: SoLAR. Edited by Lang, C., Siemens, G., Wise, A., & Gasevic, D. pp. 49-57.
- Rubel, A. and Jones, K.M.L. (2016) Student privacy in learning analytics: An information ethics perspective. *The Information Society*. 32,2 (Mar 2016), 143-159. DOI: <https://doi.org/10.1080/01972243.2016.1130502>
- Tsai, Y.-S., Moreno-Marcos, K., Tammets, K., Kollom, K., & Gašević. (2018). SHEILA policy framework: informing institutional strategies and policy processes of learning analytics. In *LAK'18: International Conference on Learning Analytics and Knowledge*, March 7-9, 2018, Sydney, NSW, Australia. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3170358.3170367>

Automated Essay Scoring in Foreign Language Students Based on Deep Contextualised Word Representations

Bojana Ranković

EPFL (Switzerland)

bojana.rankovic@epfl.ch

Sarah Smirnow

University of Zurich (Switzerland)

sarah.smirnow@ibe.uzh.ch

Martin Jaggi

EPFL (Switzerland)

martin.jaggi@epfl.ch

Martin J. Tomasik

University of Zurich (Switzerland) and University of Witten-Herdecke (Germany)

martin.tomasik@ibe.uzh.ch

ABSTRACT: We introduce a method for automated grading of handwritten essays written by foreign language learners of French. The handwriting recognition system allows digitising the essays for further processing and functions at a low character error rate. The transcriptions are then vectorised using embeddings from state-of-the-art pre-trained natural language processing models. On top of the extracted word-level features, a deep recurrent network was trained for grade predictions for essays, using the nine different grading criteria as target variables. Scores on these criteria were previously obtained from human expert raters for more than 6'000 student essays. We present preliminary findings on prediction accuracy and discuss possible future developments and applications of the system.

Keywords: Assessing writing, automated grading, handwriting recognition, natural language processing

1 BACKGROUND

Providing students, teachers and schools with objective and reliable information about students' writing competency as well as providing an evaluation related to their reference groups has become particularly relevant. This is in response to the so-called PISA shock in 2000, when, against all expectations, Switzerland was ranked just above average (e.g. Buschor, Gilomen, & McClusky, 2003). This disappointing result spurred efforts to improve Switzerland's educational system. One of the measures implemented was regular assessments of students' competencies. In our presentation, we will focus on the assessment of writing competencies in an initiative of four German-speaking cantons. More specifically, we will outline the assessment and scoring of handwritten, paper-based writing assignments in French as a foreign language as part of a set of compulsory standardised, large-scale assessments that are administered in grade eight ($N > 12'000$). We will explore digital automated scoring to better understand text features that differentiate between competence levels and to define a model for evaluating texts to a specific prompt to support human raters.

2 WRITING ASSESSMENTS

Each test consists of two open writing tasks that require students to write different types of texts (e.g. letters, messages, stories, reports) in French. These texts are analytically scored by expert readers, using a standardised grid in which different elements of the text are evaluated according to verbal gradations (Weigle, 2002). Essays are rated against nine criteria within two dimensions. The content dimension is operationalised within five criteria (task fulfillment, comprehensibility, creativity, coherence and greetings in the case of letter writing), and the language dimension is operationalised within four criteria (syntax, linguistic range, and grammatical and orthographic competence). The selection of these criteria is guided by the communicative and linguistic abilities reflected in the written product (e.g. CEFR, 2001). A major challenge in testing writing competences with open tasks is the time and expense needed for scoring (Page, 1968). Training for raters takes place during the first two days of a scoring period, and various procedures are used throughout the whole period to ensure accurate and consistent scoring. After this extensive training, raters achieve interrater consistencies of $.96 < r_{ICC} < .97$ (computed as one-way, multiple rater consistency), according to McGraw and Wong (1996), depending on the respective dimension.

3 TECHNOLOGICAL INNOVATION

Our proposed system consists of two phases. The first is digital handwritten text recognition (HTR) of more than 6'000 essays, and the second is the prediction of the annotated scores and the highlighting of interpretable features (such as keywords or other patterns) that contribute and correlate with each score or level of competence.

3.1 Handwriting recognition

In an interdisciplinary effort, the handwritten texts were digitised in order to apply natural language processing (NLP) techniques to analyse the prescored essays on each dimension. The HTR system used a neural network architecture with convolutional and recurrent layers to achieve a reliable performance of 8% character error rate on average. Despite the high heterogeneity in students' handwriting, this error rate is comparable to state-of-the-art methods in the field (e.g. Slimane, Mazzei, Topalov, Verzi, & Kaplan, 2017). Taken together, this initial part of the study served to transcribe handwritten essays into a digitised form for further evaluation using NLP.

3.2 Essay Feature Representation

In the main part of the study, the resulting transcribed texts from the HTR phase were vectorised using embeddings from state-of-the-art pretrained NLP models (BERT; Devlin et al., 2018). The model was chosen for being capable of modeling semantic and syntactic characteristics of word use while also being able to distinguish between different linguistic contexts and thereby successfully modeling polysemy. We employed a model pre-trained on French (CamemBERT; Martin et al., 2019). On top of the extracted word-level features, a deep recurrent network was trained for predicting grades on entire essays, using the nine different grading criteria as target variables. The initial training was carried out using a subset of 100 essays split into test and train examples through 5-fold cross-validation.

3.3 Representation Depth and Scoring Predictability

The results are satisfying on all of the nine criteria showing that the system can be used to support human raters. An example is presented in Table 1. Here we can also observe how different encoding layers of BERT architecture influence prediction results. The numbers show mean squared error evaluated on separate cross-validation folds. Lower-level encoding layers are able to capture syntactic rules which, in turn, enables the model to achieve lower mean squared error using these layers. In the case of originality, however, higher-level layers capture semantics and present a better choice for this content dimension.

Criteria	Bert encoding layers		
	Low-level	Middle-level	Top-level
Syntax	0.39±0.02	0.35±0.01	0.45±0.04
Originality	0.55±0.08	0.53±0.03	0.49±0.04

Table 1: MSE results for different choice of layers

4 SUMMARY AND OUTLOOK

In our presentation, we will discuss the detailed prediction accuracies in all criteria, as well as the features that were identified as the most predictive for human ratings and discuss future developments of the system. The reliable HTR system, in combination with the automated evaluation of an essay for different criteria makes it possible to develop systems that not only grade students' essays but also provide formative feedback, thereby helping to significantly improve writing style (e.g. Wingate, 2010, Malik et al., 2019). We can assume that such systems would be particularly effective if embedded in more comprehensive feedback or tutoring systems that collect and process broader data on students' competencies. One could also imagine teachers using end-to-end systems on handheld devices that can support them in evaluation and feedback. Future research is needed to test how well the models generalise to (a) different essay topics and (b) to different languages.

REFERENCES

- Buschor, E., Gilomen, H., & McCluskey, H. (2003). *PISA 2000: Synthese und Empfehlungen*. Neuchâtel, Switzerland: Bundesamt für Statistik.
- CEFR, Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching and assessment*. Cambridge, England: Cambridge University Press.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Martin, L., Muller, B., Suárez, P. J. O., Dupont, Y., Romary, L., de la Clergerie, É. V., ... & Sagot, B. (2019). CamemBERT: a Tasty French Language Model. *arXiv preprint arXiv:1911.03894*.
- Page, E. B. (1968). The use of computers in analyzing student essays. *International Review of Education*, 14, 201–225.
- Slimane, F., Mazzei, A., Topalov, O., Verzi, G., & Kaplan, F. (2017). A web-based tool for segmentation and automatic transcription of historical documents. *Proceedings of the 2017 International Joint Conference on Neural Networks*, 2730–2737.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge, UK: Cambridge University Press.
- Wingate, U. (2010). The impact of formative feedback on the development of academic writing. *Assessment and Evaluation in Higher Education*, 35, 519–533.

Scaffolding Teachers Sense Making for Classroom Equity using Visual Analytics

Ali Raza, William R. Penuel, Jennifer Jacobs, Tamara Sumner
Institute of Cognitive Science, University of Colorado Boulder
{a.raza, william.penuel, jennifer.jacobs, sumner}@colorado.edu

ABSTRACT: Creating equitable learning environments is a critical issue for teachers with diverse classrooms, whose students come from a wide range of backgrounds and cultures. Visual analytic tools, coupled with the surveys of student experience, have a promising role to play in helping teachers notice patterns in inequitable participation in classroom activities. We engaged seven middle school science teachers in a design process to create a visual analytic tool - the Student Electronic Exit Ticket (SEET) system – to facilitate their sense-making of classroom equity data and help them to identify possible classroom inequities.

Keywords: Equity, Visual Analytics, Design, Sense making

1 INTRODUCTION

A key goal of teaching today is enabling all students to contribute equitably to knowledge-building activities focused on explaining real-world phenomena (Penuel & Watkins, 2019). We are studying how visual analytics can help teachers reflect on the degree to which all students in their classroom feel welcome and able to contribute; we define *equitable* classrooms where students' quality of experience cannot be predicted by their gender, race or home language. Our approach for collecting information from students about their classroom experiences relies on experience sampling (Larson & Csikszentmihalyi, 2014). At the end of a class period, a short survey called "Student Electronic Exit Ticket" (SEET) is deployed to assess students' experiences based on three constructs: perceived coherence of the lesson, relevance to self and community, and contributions to knowledge-building. SEET questions related to coherence ask students whether they understand how current classroom activities contribute to the larger investigations in which they are engaged. Relevance questions ask students to consider the degree to which lessons matter to the students themselves, to the class, and to the larger community. For contribution, the SEET survey asks students whether they shared their ideas in a group discussion, heard ideas shared by others, and whether others' ideas impacted their thinking. Prior research has shown that these measures are reliable indicators of differentiating classrooms with different practices and learning outcomes (Penuel et al. 2018).

1.1 Problem and Motivation

As K12 classrooms have become increasingly diverse, research has recognized the need for providing teachers with effective strategies for productively engaging diverse students and providing equitable opportunities to learn (Garcia & Guerra, 2004). Problem- and project-based approaches that connect to students' everyday lives are one strategy for engaging diverse learners (NASEM, 2018). However, students' experience of these approaches varies (Penuel et al. 2016). We conjecture that capturing students' experience data with a visual analytic tool can help teachers notice inequities in student experience and reflect on their instructional practices and track how changes in their practices impact their students. Towards this end, we have conducted a series of studies with middle school teachers to examine how different visual analytic feedback displays support noticing, reflecting on, and making sense of potential inequities in their classroom.

2 ADOPTED APPROACH

Figure 1 illustrates the teacher's ideal workflow when using the SEET system. A teacher launches the survey measure for that class and students complete the survey. Then the student data from the survey is visualized on a teacher facing dashboard. The teacher then engages in sense-making to resolve a perceived gap in knowledge on classroom inequities from the dashboard and finally identify action based on the student experience of the classroom (Bertrand & Marsh, 2015).

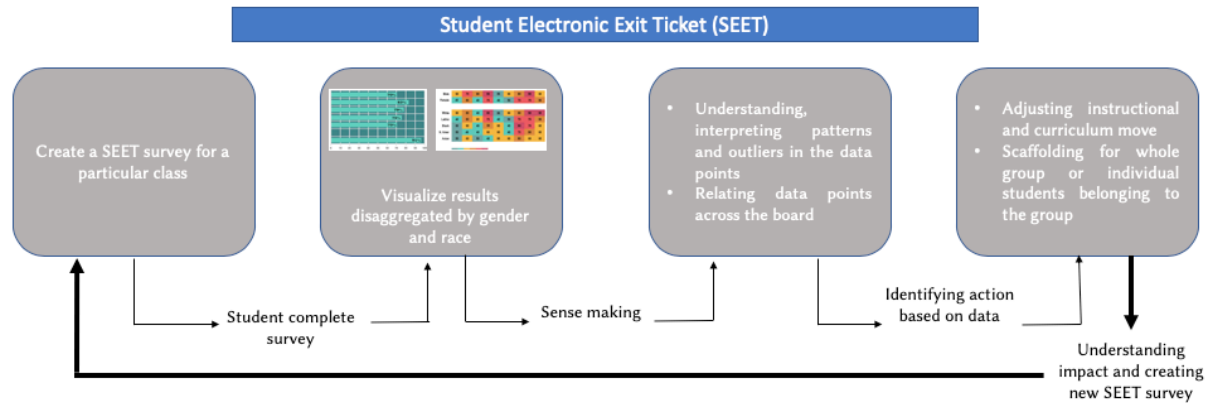


Figure 1: SEET System Architecture

3 FINDINGS

Researchers from the Learning analytics (LA) community emphasized the importance of the human-centered methods being incorporated in the field of LA (Buckingham Shum et al. 2019). We adopted a human-centered approach by involving teachers in providing feedback on the SEET system. We presented them with a total of thirty different visualizations of SEET data that displayed data on the gender and race disaggregation created from a real dataset having significant classroom inequities of experience. We had them think aloud as they made sense of data, to help visualizations that facilitated noticing of inequities. All teachers included in the design process taught science and engineering at the middle school level and were using a problem-based approach to teaching.

Table 1 below details the preferred visualizations by different teachers in both iterations. All names used are pseudonyms. During iteration 1, the first three teachers selected visualizations (Horizontal stacked bar, connected scatterplot) as it supported them in the sense-making of data. So, we switched to different visualizations with the last two teachers in iteration 1. These two teachers selected these visualizations (Heat map, Bubble chart, line chart). Starting the second iteration with two new science teachers, we selected preferred visualizations from the first iteration along with some new visualizations for conforming to our selection on visualizations that are aligned with the sense-making of teachers. From this iteration, the two teachers selected visualizations (Horizontal Histogram, Heat map, Connected scatterplot). We finalized 'Horizontal Histogram', 'Heat map' disaggregating data of gender and race, and 'connected scatterplot' for overtime investigation.

Table 1: Visualizations preferred in sense-making

Participant	Preferred Visualization	Visual Cue
Yarosh~	(1) Horizontal stacked bar**, (2) Connected scatterplot	(1) Size, (2) Space
Meer~	(1) Horizontal stacked bar**, (2) Connected scatterplot	(1) Size, (2) Space
Tim~	(1) Horizontal stacked bar**, (2) Connected scatterplot	(1) Size, (2) Space

Becky~	(1) Heat map**, (2) Bubble chart**, (3) Connected scatterplot	(1) Color, (2) Size, (3) Space
John~	(1) Heat map**, (2) Bubble chart**, (3) Connected scatterplot	(1) Color, (2) Size, (3) Space
Jack~~	(1) Horizontal Histogram, (2) Overtime, (3) Heat map	(1) Size, (2) Space, (3) Color
Alan~~	(1) Horizontal Histogram, (2) Overtime, (3) Heat map	(1) Size, (2) Space, (3) Color

~Iteration 1, ~~ Iteration 2, *Overall, **disaggregated, ***Overall and disaggregated

A vignette design moment appeared during the first iteration of the study, we used the stacked histogram to visualize disaggregated data on race and gender. Teachers provided insight to display the only percentage 'yes' to a question, rather than displaying 'no', 'I don't know' along with it. We iterated with our design and in the second iteration, we adopted a horizontal histogram displaying only percentage 'yes', that lead to an improved sense-making for the teachers. Based on our design study findings, we incorporated Histogram, Heat map, and Connected scatterplots for visualizing student experience data on 'coherence', 'relevance', and 'contribution' constructs. We selected these visualizations based on a combination of two factors: ease of interpretation, and how well they facilitated noticing of inequities in data.

4 CONCLUSIONS

Our research with the science teachers facilitated the selection of the visual feedback displays appropriate for the sense-making of teachers. We are not implying these displays as the *sine qua non* for future designers of the visual dashboard, instead, more work is warranted to establish concrete measures. Teachers also provided insights on how particular displays could help them in understanding classroom inequities. Further, we aim to test the use of these displays with teachers in the future who are testing strategies intended to create more equitable experiences for students in classrooms.

REFERENCES

- Bertrand, M., & Marsh, J. A. (2015). Teachers' sensemaking of data and implications for equity. *American Educational Research Journal*, 52(5), 861-893.
- Buckingham Shum, S., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-centred learning analytics. *Journal of Learning Analytics*, 6(2), 1-9.
- Garcia, S. B., & Guerra, P. L. (2004). Deconstructing deficit thinking: Working with educators to create more equitable learning environments. *Education and urban society*, 36(2), 150-168.
- Larson, R., & Csikszentmihalyi, M. (2014). The experience sampling method. In *Flow and the foundations of positive psychology* (pp. 21-34). Springer, Dordrecht.
- National Academies of Sciences, Engineering, and Medicine. (2018). How people learn II: Learners, contexts, and cultures. *National Academies Press*.
- Penuel, W. R., Van Horne, K., Jacobs, J. J., & Turner, M. (2018). Developing a validity argument for practical measures of student experience in project-based science classrooms. In *Annual Meeting of the American Educational Research Association*, New York, NY.
- Penuel, W. R., & Watkins, D. A. (2019). Assessment to Promote Equity and Epistemic Justice: A Use-Case of a Research-Practice Partnership in Science Education. *The ANNALS of the American Academy of Political and Social Science*, 683(1), 201-216.
- Penuel, W., Van Horne, K., Severance, S., Quigley, D., & Sumner, T. (2016). Students' responses to curricular activities as indicator of coherence in project-based science. Singapore: *International Society of the Learning Sciences*.

A note on Automatic Grading of Short Answers and Providing Feedback

Neslihan Süzen, Alexander N. Gorban, Jeremy Levesley and Evgeny M. Mirkes

University of Leicester, Leicester LE1 7RH, UK

ns433@leicester.ac.uk

ABSTRACT: In this study, we focus on automatic scoring of short answer questions that are typical in the UK GCSE system, and providing useful feedback to students. Standard data mining techniques were applied to the set of students' answers in order to measure similarity between student answers and the model answer. We evaluated the relation between these similarities and marks given by scorers. We also performed a clustering algorithm that groups students' answers into clusters and showed that students who are awarded the same or similar scores are grouped together. This study aims to design a mathematical model to predict marks based on the similarity defined. We argue that computational methods be used to enhance the reliability of human scoring and not replace it. Human are required to calibrate the system, and to deal with challenging situations.

Keywords: Natural Language Processing, Information Extraction, Automatic Grading, learning Process, machine Learning, Text Mining, Clustering, Assessing Student learning

1 INTRODUCTION

Manual assessment of short answer and essay questions is more difficult and takes considerable time as it requires textual understanding and evaluation (McDaniel, Anderson, Derbish, & Morrisette, 2007; Mohler, Bunescu, & Mihalcea, 2011; Reynolds, Livingston, Willson, & Willson, 2010). It also leads to inconsistency in grades from two graders, as they must infer the meaning in answers. In this study, we present our approaches to automatic marking and providing feedback for short answers – usually no more than 30-40 words. The initial assumption is: marks highly depend on words that students used which also appear in model answer. We hypothesizes that students having lack of knowledge are not able to use all words (or synonyms) in the model answer, and marks can be predicted by distance (or similarity) between the model answer and student answer. We develop a model to predict marks using this distance. This marking is done under the assumption that the text is an attempt to answer the question in English. This can be checked without subject-specific expertise. The second focus of this study is automatic feedback mechanisms. Our approach is based on clustering students' answers into groups to discover natural groups of similar answers. We can provide teachers with information about common answers since students generally answer questions in similar way. A system can be built by selecting prototype answer(s) from clusters, providing marks and feedback for the prototype and then assigning them to the entire cluster. Such a system can significantly reduce time in manual assessment. We do not aim to remove the human from assessment process, but to improve consistency in marking and feedback and allow human to apply judgement in difficult cases. The dataset used consists of students' answers and a model answer for each question from a computer science class at the University of North Texas. Answers are graded by two human judges using marks from 0 to 5. We demonstrate our approaches with the

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

first question in data due to its representativeness of average short answer length (1 sentence in model answer)¹. We define the *model vocabulary* as unique words in the model answer. The model answer for the first question is “**To simulate the behavior of portions of the desired software product.**” and the model vocabulary contains **simulate**, **behavior**, **portion**, **desire**, **software**, and **product**. In clustering based approach, we used k-means clustering using Euclidean distance (Luo, Li, & Chung, 2009; Suzen, 2019).

2 MARKING AND FEEDBACK MECHANISMS

We suspect that those students graded at the lowest score did not use appropriate terminology and those graded at the highest score used appropriate terminology. In between, they may use some words requires but the answer is not very appropriate. Therefore, we turn to looking at natural clusters of answers to discover hidden patterns. Figure 1 shows three clusters found, grades of answers and frequently used words in clusters. Clusters *Excellent* and *Weak* are well separated in terms of scores, which are 5 and 2. We did not identify any scoring rule in cluster *Mixed* where there is also discrepancy in scores graded by two teachers. This shows that teachers find it challenging to score responses in this cluster. Students in *Excellent* cluster used all words of the model vocabulary, while there is no model word in the cluster *Weak*. Hence, we are able to easily separate clusters when students use all or none of words from model vocabulary. A few model words appear in cluster *Mixed* and synonyms of ‘part’ and ‘final’ (portion, desired). Marks change with words used in this cluster. A better knowledge of acceptable vocabulary is needed to cluster more effectively.

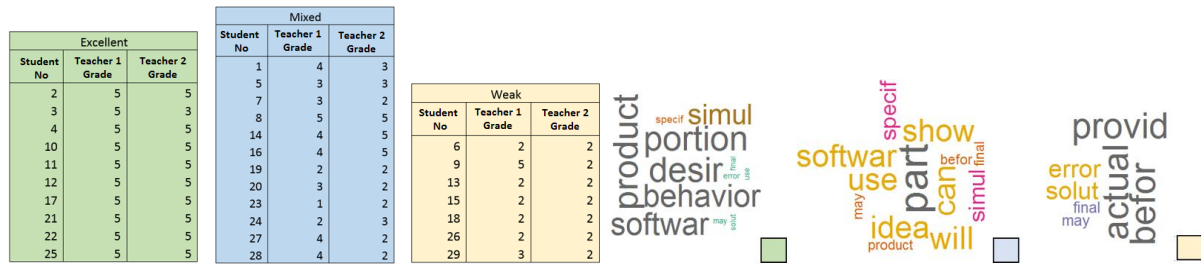


Figure 1: (a) Clusters along with scores by two human judges and (b) frequent words in each cluster.

We now consider how marks depend on the model vocabulary used by students. The number of model words (n) that a student used is counted and the Hamming distance between a student answer r and the model answer m is calculated as $h(r,m)=6-n$. Scores decrease as the distance increases. The Pearson correlation between the distances and marks given is -0.81 and -0.83 for two teachers. Thus, teacher assessment highly depends on how many of model words students used. We then hypothesis that this distance is a strong indicator of the mark of a student, which makes automated scoring possible. A mathematical model (MM) that predicts marks is created and evaluated (Suzen, Gorban, Levesley, & Mirkes, 2018). The predictor function that models this relationship is $y = \beta_0 + \beta_1 h + \beta_2 h^2$ where β_0 , β_1 and β_2 are parameters, and h is the distance (0, 2, 4, 5 or 6 in this question). The average of two teachers' marks is used as the actual value of the dependent variable, so-called TM . The estimated parameters for MM are $\beta_0=4.9108$, $\beta_1=-0.0058$ and $\beta_2=3.4236$,

¹ The dataset used was downloaded from the archive hosted at the URL <http://lit.csci.unt.edu/index.php/Downloads>. Available at https://github.com/dbbrandt/short_answer_grading_capstone_project/tree/master/data/source_data (accessed 16 December 2019).

and MSE for MM is 0.17. We calculated the deviation of teacher grades from the average mark and found that MSE (TM) is 0.25. Thus, teachers' marks diverge from their average more than does the MM. Figure 2 shows the predicted and actual marks. The mark is decreasing function of distance in general. Both models show qualitatively the same behavior. The disagreement between teachers increases significantly with the distance until we regain agreement for poor answers. MM outperforms TM for distance 5 and 6. Therefore, we conclude that MM is no worse than human marking. The model lies close to the average of grades of two markers; and appropriate for some questions than others, but those questions that the algorithm finds hard to mark are also difficult for humans (Suzen, Gorban, Levesley, & Mirkes, 2018).

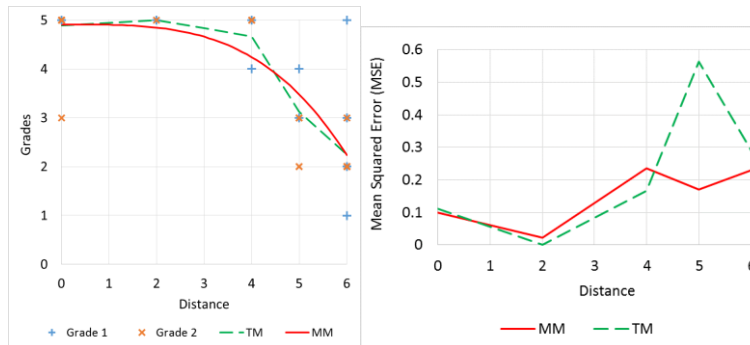


Figure 2: (a) Actual marks from two human judges with TM and predicted marks by MM with distances from the model answer; (b) MSE for each distance for MM and TM.

We conclude that our methodology works well where a model vocabulary can be clearly identified. This can be automated in some cases, but may require human input. The strong correlation of grades and the distance suggests that marks can be predicted by the distance with high accuracy. If a large number of responses are being graded, it is reasonable that a human would move towards pattern recognition via key words rather than reading for meaning. This allows time saving for scoring, and to provide rapid feedback by checking words from the vocabulary. Such an automatic scoring system can provide a clear baseline where conversation about assessment and feedback can develop. It is crucial that in this age of improving artificial intelligence, that we use machines to reduce the amount of repetitive straightforward scoring, which human is poor at performing, and have people engaged in higher level, more valuable assessment.

REFERENCES

- Luo, C., Li, Y., & Chung, S. M. (2009). Text document clustering based on neighbors. *Data & Knowledge Engineering*, 68(11), 1271-1288.
- McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, 19(4-5), 494-513.
- Mohler, M., Bunescu, R., & Mihalcea, R. (2011). *Learning to grade short answer questions using semantic similarity measures and dependency graph alignments*. Paper presented at the Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1.
- Reynolds, C. R., Livingston, R. B., Willson, V. L., & Willson, V. (2010). *Measurement and assessment in education*: Pearson Education International Upper Saddle River.
- Suzen, N., Gorban, A., Levesley, J., & Mirkes, E. (2018). Automatic Short Answer Grading and Feedback Using Text Mining Methods. arXiv:1807.10543 [cs.CL] (*Procedia Computer Science*, accepted).
- Suzen, N. (2019). Automatic-short-answer-grading-and-feedback-project. Available from <https://github.com/neslihansuzen/Automatic-short-answer-grading-and-feedback-project>

Chatbots – An Opportunity for Individual Assistance in Education

Sebastian Wollny

DIPF | Leibniz Institute for Research and Information in Education
Wollny@dipf.de

Jan Schneider

DIPF | Leibniz Institute for Research and Information in Education
Schneider.Jan@dipf.de

Marc Rittberger

DIPF | Leibniz Institute for Research and Information in Education
Rittberger@dipf.de

Hendrik Drachsler

DIPF | Leibniz Institute for Research and Information in Education
Drachsler@dipf.de

ABSTRACT: The communication behavior of us humans has changed greatly in recent years. The ubiquitous presence of digital media, networked devices and technology has created new types of conversations, discussions and information sharing. However, the resulting possibilities for designing lessons and supporting learners have not yet been fully exploited. Especially in the individual support of learners, current forms of education often reach the limits of their resources. Promising technologies in this respect are Chatbots. They conduct conversations automatically in natural language and can be addressed by learners in case of problems or comprehension questions. By reproducing information from predefined knowledge structures, Chatbots enable the development of digital education assistants that respond individually to learner inputs.

Keywords: Chatbots, Digital Assistants, Artificial Intelligence, Personalization, Recommender Systems, Natural Language Processing

1 INTRODUCTION

Chatbots allow natural communication between humans and machines. In addition, they are able to recognize intentions and react according to a set of rules defined by curated example conversations. Since the invention of ELIZA in 1966 (Weizenbaum, 1966), it has been possible to develop Chatbots, which are capable of small talk and simple forms of information transfer. The underlying technology has evolved over the years to include advanced language models and Natural Language Understanding (NLU) modules. Nowadays, Chatbots are one of the main potential applications of Artificial Intelligence (AI) in education. They can be used to access a variety of data sources, help with orientation or independently handle entire process chains without human intervention. Especially due to the increasing use of digital assistants, Chatbots are becoming more and more tangible for a large number of people.

The advantages of Chatbots could be helpful in order to make better use of the limited resources of teachers and to improve the amount of individual learner feedback.

In this paper, we present a Chatbot concept, which describes a solution for Self-Regulated Learning (SRL). It consists of an educational and a technical perspective. The educational perspective deals with the positioning of Chatbots in learning theory, whereas the technical perspective provides a technical implementation guideline.

2 EDUCATIONAL PERSPECTIVE

The learner-centered approach of SRL transfers responsibility to the learner. The common concept of SRL (Zimmerman & Moylan) structures the necessary actions in three steps: In the forethought phase, learners set goals that they want to achieve. In the performance phase, learners observe themselves as they continue to work towards achieving their goals. Finally, a SRL cycle is closed by the Self-Reflection Phase, whose results are used in the next Forethought Phase.

Considering possibilities of using Chatbots in these processes, Chatbots have the potential to support learners at all three phases, while retaining control over the learning cycle as the learner. Chatbots could improve SRL processes to be more efficient or effective by supporting or assisting learners for example through triggering planning actions, supporting self-monitoring or asking reflective questions (Wollny, Schneider, Rittberger, & Drachsler). This leads to an extended learning cycle, which is shown in Figure 1 and will be evaluated later in our studies.

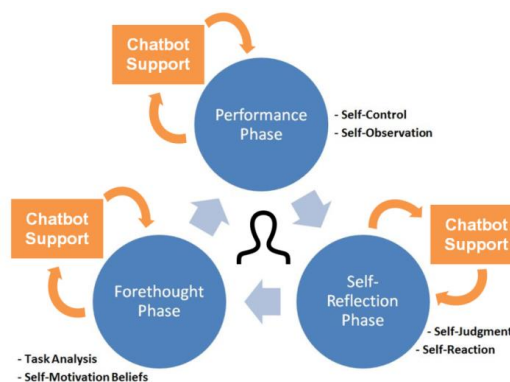


Figure 1: Chatbots in context of Self-Regulated Learning

3 TECHNICAL PERSPECTIVE

Writing text messages is nowadays the preferred way of communication for many people at the age of student (Rideout & Robb, 2018). Integrating automated messages into messaging apps or services is in many cases possible through APIs. The underlying technology of Chatbots thus allows us to give easily accessible and individual learner feedback in the same way learners communicate with each other around the clock.

One major feature of Chatbot implementations is the automated answering of frequently asked questions. By defining appropriate domain knowledge, Chatbots could help in our case with issues

such as exam anxiety or procrastination. Moreover, answering general questions on how to structure and plan the daily life could help learners with keeping control over their progress.

For helping learners with advice in their performance phase, Learning Analytics data is required that provides insights into the actions of learners. The personal learner data can be brought together in an individual learner profile, which is fed from LMS data and answers from responses to directly asked questions about the learner. Looking at the Chatbot system from a technical perspective, it bridges the gap between already used messaging apps, learner profiles and domain knowledge.

This implementation of a Chatbot has to take the general structure of these systems into consideration. Chatbots are technically divided in two components (Figure 2): A Natural Language Understanding (NLU) Unit and Storyline Unit. The NLU Unit is responsible for extracting or classifying the learner's intention, while the Storyline Unit provides a conversation guide. Both systems communicate with each other via structured data channels and are enhanced by recorded and annotated learner-chatbot conversations.

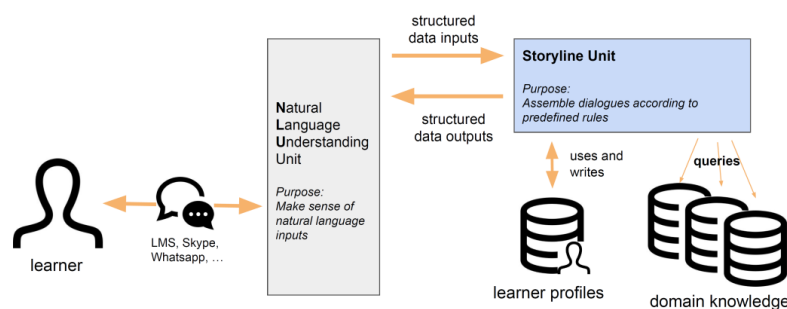


Figure 1: Technical implementation concept of a digital learning assistant Chatbot

4 OUTLOOK

The Chatbot concept proposed in this paper should support self-regulated learners. In our next step, we want to investigate in a case study, if Chatbot feedback has an effect on how students perform in SRL. Subsequently, we want to compare SRL technologies in a user study.

REFERENCES

- Rideout, V., & Robb, M. B. (2018). *Social media, social life: Teens reveal their experiences*. San Francisco, CA: Common Sense Media.
- Weizenbaum, J. (1966). ELIZA---a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45.
- Wollny, S., Schneider, J., Rittberger, M., & Drachsler, H. Am I planning smart?—Analyzing student goals. *Companion Proceedings of the 9th International Learning Analytics and Knowledge Conference (LAK'19)*, 261–266.
- Zimmerman, B. J., & Moylan, A. R. Self-regulation: Where metacognition and motivation intersect. In *The educational psychology series. Handbook of metacognition in education 2009* (pp. 311–328).

De-identification is not enough to guarantee student privacy: De-anonymizing personal information from basic logs

Elad Yacobson

Weizmann Institute of Science
elad.yacobson@weizmann.ac.il

Orly Fuhrman

The Center for Educational Technology
orlyf@cet.ac.il

Sara HersHKovitz

The Center for Educational Technology
sarah@cet.ac.il

Giora Alexandron

Weizmann Institute of Science
giora.alexandron@weizmann.ac.il

ABSTRACT: Learning Analytics have the potential to significantly improve teaching and learning. However, if student data is reused and transferred, protecting privacy becomes a major concern. A common belief among education experts is that de-identification is enough to protect student privacy. However, this is not the case. Our study demonstrates how several demographic characteristics can be mined from de-identified student log files coming from a web-based tutor. The logged events include only two attributes: the time and correctness of students' attempts. We show how we can use these attributes to identify the physical classes and schools, by intersecting temporal patterns with publicly available school data. Our goal is to convey the message that i) de-identification alone does not guarantee student privacy; ii) removing data attributes that can be used to re-identify learners would render the data useless for learning analytics. This emphasizes that trust, backed-up by appropriate legal agreements, should be the cornerstone of data-sharing policies.

Keywords: Learning Analytics, Privacy, Re-identification

1 INTRODUCTION

Learning Analytics (LA) have the potential to improve teaching and learning using rigorous, data-driven methods (Siemens, 2013). In many cases, LA involve the transfer of data between the data owner and the data processor, yielding tension between privacy and openness (Daries et al., 2014). A common approach for protecting privacy is de-identification, but the stricter the de-identification, the greater the negative effect on the ultimate analysis (Khalil & Ebner, 2016). In addition, de-identified data have the risk of being re-identified, for example by linking with other datasets (Sweeney, 2000).

Our goal is to examine the tension between usefulness and anonymity, and provide evidence that cleaning the data from any attribute that can be used to re-identify learners will render the data useless for LA. The rationale is that even the most basic interaction patterns create a 'personal' footprint, and that combining this with contextual information, which is also required for making LA meaningful, can reveal various properties of the learners.

Our focus is on LA in K-12. Specifically, the research question studied is whether we can re-identify the classes and schools that worked with an Intelligent Tutoring System, based on logs that contain only two attributes of students' interaction: the time of each attempt, and whether it was correct or not.

While we do not de-anonymize personal information, the information that we do mine may be sensitive, for example if a school does not wish to reveal student achievements. In addition, our approach could have been taken further to re-identify more sensitive information. However, we deliberately decided to avoid trying to re-identify *individual* learners.

2 METHODOLOGY

Experimental Setup. The experimental setup is based on a web-based reading comprehension tutor. The tutor was used in sixteen classes by ~500 fifth grade students. The students typically worked with the tutor for one or two hours per week, within regular school hours, for about two months. The tutor contains twenty learning units, each composed of a text, followed by a dozen questions. Clickstream data were collected by the tutor, but for this study we used only two attributes of these data: The time and correctness of students' attempts.

Process. We used an unsupervised approach. The process was composed of the following steps:

Step 1 – Building time intervals: We built, per student s , a list I_s which contains the time intervals in which the student worked with the system. Each interval within I_s contains events that are less than sixty minutes apart, and its boundaries are defined by the time of the first and last event.

Step 2 – User-user similarity matrix: We used a user-based collaborative filtering approach. First, we computed the Jaccard similarity index between each pair of students. The Jaccard index is defined as $(A, B) = \frac{|A \cap B|}{|A \cup B|}$. In our case, A and B are time intervals I_i, I_j . $|I_i \cap I_j|$ = the number of overlapping intervals, where two intervals are considered as overlapping if at least half of each is contained in the other. The result is a user-user similarity matrix M , where $M[i, j] = \text{Jaccard}(I_i, I_j)$. M is symmetric.

Step 3 – Clustering the students: We ran K-Means on M to cluster the students into classes. The number of classes was determined using Gap statistic (Tibshirani et al., 2001).

Step 4 – identify class attributes: After clustering the students into classes, we examined students' achievements in the tutor, identified the weekly slots in which each class tended to work, whether there were deviations from this weekly routine, the date of the first and last interaction of each class, etc.

Step 5 – de-anonymization: We intersected the patterns discovered in Step 4 with publicly available information to identify the physical location of the classes and schools.

3 RESULTS

3.1 Group students into classes

As described in Subsection 2.2, our method used a user-based collaborative-filtering approach with Jaccard index and K-Means clustering. Gap statistic yielded fourteen classes, but when constraining classes to contain <40 students, it yielded sixteen classes, which is the true value. We measured the goodness-of-fit of the clustering with Adjusted Rand Index (ARI) (Hubert & Arabie, 1985). The ARI for the K-Means clustering was 0.977. We compared the collaborative-filtering approach to a Union-Find (connected components) approach: Each two students i, j are considered as belonging to the same class if at least 30% (obtained empirically) of their time intervals intersect. If yes, the classes of i and j are merged. The ARI for the Union-Find algorithm was 0.669.

3.2 Identifying Demographic Properties

After clustering the students into classes, we examined students' achievements in the tutor (since the compatibility between the actual classes and the clusters yielded by the algorithm was almost 100% accurate, we hereafter use the term "class" when referring to a certain cluster). We found one class with a significantly higher performance, and one with considerably lower performance, suggesting that these are special classes for gifted children and children with learning disabilities.

Both classes started and finished using the tutor at the exact same dates, and used the tutor on similar hours and similar days of the week. Since the start and end dates of the other classes were different, we hypothesized that both classes belong to the same school. A web search yielded that there are only three schools in the country that have both types of special classes (gifted children and learning disabilities) within the same school. Next, we looked for patterns in the time slots used by these two classes, and whether there were deviations from this regular pattern. This yielded that the (presumably) gifted children class used the tutor on a certain date at irregular hours.

We then checked the schools' webpage, along with the relevant towns' local newspapers webpage, to see if there was any special event on that specific date which could explain this irregularity. We found that one of the above-mentioned schools was on a field trip that morning, suggesting that the gifted class belonged to that school. **Checking with the company, confirmed the identification was correct.**

4 DISCUSSION AND CONCLUSIONS

The context and settings in which students use online learning environments is naturally reflected in student data. For example, if a class learns a certain topic together from an online learning application, the logs of the students will include interactions with certain items during a bounded time window. Our results demonstrate that even the most basic attributes of such online learning interaction – time and correctness of learners' attempts – create patterns that can be used to reveal student properties such as schools and classes. Since time and correctness are the most basic information required by LA algorithms (e.g., adaptive learning), this means that such algorithms require data that contain sensitive information, even if *direct identifiers* (e.g., user name, email, ip) are removed.

LA have a lot of potential, but are often too expensive to develop in-house. This drives a growing market of third-party LA providers, which is a good sign of a maturing field. However, our results underline that sharing behavioral data with such providers should be regarded as carrying a potential risk to student privacy, and thus should be done with trusted partners, and backed by appropriate legal agreements, which for example ban linking application data with external data sources.

REFERENCES

- Daries et al (2014) Privacy, anonymity, and big data in the social sciences. *Commun. ACM* 57 (9).
- Hubert, L., & Arabie, P. (1985) Comparing partitions. *Journal of classification*, 2(1).
- Khalil M., & Ebner, M. (2016) De-identification in learning analytics. *Journal of Learning Analytics* 3(1)
- Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, 57(10).
- Sweeney, L. (2000). Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000.
- Tibshirani, R., Walther, G., & Hastie, T. (2001), Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society*, 63(2)

What Do You Get from Replies? Causal Estimates of Peer Effects in Online Discussion Forums

Renzhe Yu

University of California, Irvine
renzhey@uci.edu

Di Xu

University of California, Irvine
dix3@uci.edu

ABSTRACT: Discussion forums, as a major tool to foster social interaction in online learning environments, can theoretically promote learning via increased social presence or collaborative knowledge construction. Due to the reflective nature of social interaction, it is empirically challenging to separate the *influence from peers* from *students' own abilities* which both contribute to engagement in discussion forums. As such, accurately estimating the benefits of online discussion becomes an important methodological problem in its own right. In this study, we address this problem in the context of five fully online courses at a public university. Utilizing random group assignment and fine-grained posting records, we estimate the causal effect of receiving peer responses on individual learning outcomes. More specifically, we differentiate the effects of *quantity* and *quality* of peer responses. According to our results, quantity matters but quality does not, suggesting that peer discussion might promote learning through social mechanisms but not cognitive channels. We further discuss the implications for designing interactive features in online courses.

Keywords: Social interaction; Online discussion; Causal inference; Peer effects

1 INTRODUCTION

Social interaction features have been commonplace in online course designs to make up the lack of face-to-face communication, and discussion forums remains one of the most widely employed features due to its ease of deployment and use for both instructor and students (Balaji & Chakrabarti, 2010). Theoretically, social interaction through discussion forums can benefit learning through two mechanisms. For one thing, effective student-student interaction can promote social presence, i.e. ability to project oneself socially into a community of “real people”, and further increase motivation and engagement (Richardson, Maeda, Lv, & Caskurlu, 2017). Likewise, students can share their own ideas and read and reflect on each other's thoughts in online discussions. Through this process, they learn from peers and build knowledge collectively (Stahl, Koschmann, & Suthers, 2014).

Research on online learning has empirically examined the relationship between online peer discussion and learning outcomes. While in many cases results are positive (e.g., Kent, Laslo, & Rafaeli, 2016), a major limitation is that these studies measure peer interaction using some metrics of individual engagement in the forum (e.g., number of posts authored). Observed engagement of a focal student is a function of both her own dispositions and the influence from her peers (a.k.a., peer effects). Moreover, peer effects are reflective: every student is affected by the peers who respond to her while exerting her influence on others (Bettinger, Liu, & Loeb, 2016). These complexities suggest that most engagement-based measures in existing studies cannot separate *self* and *peers* and therefore do not contribute to causal conclusions.

To understand whether peer discussion *causes* individual learning, we need to tease out peer effects that are independent of reflective discussion processes, which this study attempts to achieve. Specifically, we examine the quantity and quality of peer responses, which are respectively mapped to the foregoing mechanisms of social presence and collective knowledge construction. Thus, we can get a deeper understanding of online peer interaction through the extent to which receiving more/better peer responses affects course performance.

2 DATA & METHODS

We analyzed five offerings of two introductory online courses in public health taught in 2017 at a four-year public institution in the United States. All these offerings were taught by the same instructor and adopted the same course design. Besides one presentation, two exams and weekly lecture videos and quizzes to finish, students were required every week to author at least one post in response to a course-related prompt and one reply to a peer post in the discussion forum. The instructor randomly assigned students into closed groups of around 10 or 20 for these discussions and graded the posts every week (out of 10) based on their quality (e.g., exhibiting new ideas). We acquired institutional data, discussion data and gradebook data of all these students.

For each student, we define the quantity of peer responses as the average number of replies she received per week, and the quality of peer responses as the average score of all those replies received. We employ an instrumental variable approach to carve out variations in the quantity and quality measures that are orthogonal to the reflective discussion processes. Specifically, we use the average pairwise difference between the focal student and all her group members in early forum behavior (e.g., time lapse between actual posting time and the assigned deadline) to instrument quantity. On the other hand, we instrument quality using the average innate ability of group members who responded to the focal student, where the innate ability of each student is estimated by a dynamic regression model from the discussion score data (Bettinger et al., 2016), as represented by the time-invariant orange portion in Figure 1.

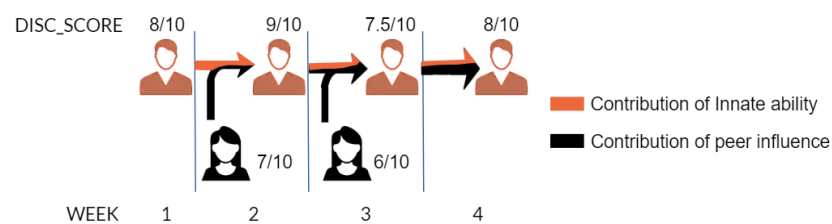


Figure 1: Estimate innate ability (in orange) from weekly discussion scores

We use final course grade (out of 100) as the learning outcome measure. After computing the instrumental variables, we run two-stage least-squares (2SLS) regressions to estimate the effect of quantity or quality of peer responses on the outcome. In these regressions, we also include instruments of the quantity and quality of focal student's own posts derived from similar approaches above, so that the causal effects of *self* and *peers* are explicitly separated.

3 RESULTS & DISCUSSION

Table 1 reports the causal estimates of peer effects, where each column is pulled from a separate 2SLS regression with a series of individual-level covariates as control. The first column shows that on

average, receiving an additional response from other group members increases the final course score by 1.14%, or one ninth of a letter grade. In the second column, we see that the quality of peer responses has no effect on individual student's learning outcome. Following our theoretical mapping, these results suggest that online discussion might afford to increase social presence but cannot effectively promote higher-order knowledge construction processes. The social mechanism might work because students received notifications when others responded to them. The failure of cognitive mechanism to work, on the other hand, may actually be explained by the fact that most students did not post more than what was required – they might “passively” participate in the discussion forum without reading and reflecting on the peer responses received. As such, standard discussion assignments might not take full advantage of peer interaction in the online world.

Table 1: Estimated effects of peer responses on final grade.
(Instrumental variable estimates)

	Quantity (count) of ...	Quality (grade) of ...
Peer responses received	1.14*** (.362)	-.0798 (.107)
Own posts	2.84*** (.372)	5.20*** (.183)
N	1,028	1,015

Note: Standard errors reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

One simple strategy to address this might be getting instructors themselves involved in discussions. Moreover, with emerging learning technologies that support a plethora of interactive features (e.g., media curation and remixing), online instructors can utilize them to create more engaging learning communities where students can augment instructional materials and learn from each other through intensive peer-to-peer activities. The sample of this study is of moderate size and somewhat homogenous course design contexts. Future work will be investigating peer interaction from a causal perspective at a larger scale and across different pedagogical contexts.

REFERENCES

- Balaji, M. S., & Chakrabarti, D. (2010). Student Interactions in Online Discussion Forum: Empirical Research from “Media Richness Theory” Perspective. *Journal of Interactive Online Learning*, 9(1).
- Bettinger, E., Liu, J., & Loeb, S. (2016). Connections Matter: How Interactive Peers Affect Students in Online College Courses. *Journal of Policy Analysis and Management*, 35(4), 932–954.
- Kent, C., Laslo, E., & Rafaeli, S. (2016). Interactivity in online discussions and learning outcomes. *Computers & Education*, 97, 116–128.
- Richardson, J. C., Maeda, Y., Lv, J., & Caskurlu, S. (2017). Social presence in relation to students' satisfaction and learning in the online environment: A meta-analysis. *Computers in Human Behavior*, 71, 402–417.
- Stahl, G., Koschmann, T., & Suthers, D. (2014). Computer-Supported Collaborative Learning. In *The Cambridge Handbook of the Learning Sciences* (2nd ed., pp. 479–500).

Predicting College Success: What Data Are Useful and for Whom?

Renzhe Yu

University of California, Irvine
renzhey@uci.edu

Qiujie Li

New York University
ql16@nyu.edu

Christian Fischer

University of Tübingen
christian.fischer@uni-tuebingen.de

Di Xu, Shayan Doroudi

University of California, Irvine
{dix3, doroudis}@uci.edu

ABSTRACT: In higher education, predictive analytics can provide actionable insights to diverse stakeholders such as administrators, instructors, and students. Separate feature sets are typically used for different layers of prediction, but little is known about the overall utility of different data sources across prediction tasks. Using data from nearly 2,000 college students at a large public university, we examined the usefulness of institutional data, learning management system (LMS) data, and survey data for accurately and fairly predicting short-term and long-term student success. We first find that institutional data most accurately predicts success among the three data sources. Also, no data source makes fair prediction: institutional data suffers group parity in false negative rate (FNR) and LMS data in false positive rate (FPR). Moreover, non-demographic minorities endure no less algorithmic biases than demographic minorities. These analyses serve to inform more cost-effective and equitable use of student data in college settings.

Keywords: Predictive analytics; Higher education; Fairness; Student data

1 RESEARCH CONTEXT

In the recent decade, predictive analytics has been playing increasingly important roles in promoting college success through data-driven applications. While more comprehensive student data are conceptually beneficial for building more accurate predictive models, collecting and managing such data is often costly for institutions. For example, learning management systems (LMS) can record hundreds of learner actions in every single minute, which requires robust and efficient data management systems. Thus, it is crucial to systematically examine the utility of different data sources for building predictive analytics-based solutions to guide administrators and policymakers on the costs and benefits of utilizing student data. However, such research is underrepresented in the literature.

This study evaluates the usefulness of three common student data sources for two representative prediction tasks by their contribution to *overall prediction performance* and to *prediction fairness across student subpopulations*. We analyzed institutional data, LMS data, and survey data, which are the most commonly used data sources to predict various measures of college success (Broadbent & Poon, 2015; Ishitani, 2006; Park et al., 2018). Typical applications utilizing these data include predicting course performance to facilitate instruction and predicting institutional outcomes (e.g., dropout) to inform academic advisors. Accordingly, we constructed two prediction targets based on course grades (short-term success) and yearly average GPA (long-term success), respectively. The focus on fairness arises from the concern that predictive models trained on the entire student population may perform systematically worse on disadvantaged subpopulations and amplify existing achievement gaps. However, empirical work on such algorithmic bias in educational settings has been limited (e.g., Hutt,

Gardner, Duckworth, & D’Mello, 2019). This study, therefore, aims to identify what combinations of student data more accurately and more fairly predict different success measures.

2 DATA AND METHODS

This study analyzed ten online introductory STEM courses taught from 2016 to 2018 in an American public university. The restriction to online courses ensured that LMS data can produce valid representations of learning behavior. We constructed three feature sets for students enrolled in these courses. Institutional features included student demographics (e.g., gender, ethnicity) and prior academic achievement (e.g., high school GPA). Click features were measures of learning behavior in Canvas LMS including the total number of clicks and total time on page over the entire course and for different time and event breakdowns. Survey features included measures of self-regulated learning skills and self-efficacy, which are essential for success in online courses (Broadbent & Poon, 2015). The survey data were collected at the beginning of each course. On the target side, we defined short-term success as whether a student’s final course grade was above the course median. Similarly, long-term success was defined as whether a student’s average GPA in the year that followed the course was above the median of her classmates. These two measures were put together with all 7 combinations of the three feature sets to form a total of 14 binary classification problems.

For each classification, logistic regression, support vector machines (SVM) and random forests were used, and the best-performing approach was chosen for further analysis. Course-level leave-one-group-out cross validation was configured to iteratively get predicted values within each course. We employed three metrics for evaluation: accuracy, false positive rate (FPR; miss out “at-risk” students) and false negative rate (FNR; “penalize” students doing well). For fairness analysis, we computed FPR disparity and FNR disparity (Saleiro et al., 2018) for each minority group and tested their statistical significance to see if the group was discriminated against by the prediction model.

3 RESULTS AND DISCUSSIONS

The final prediction model used a total of 1,871 student-by-course data points. Table 1 presents the prediction results on this full sample across different feature and target combinations. Overall, institutional data best predicts both success measures as it is almost always present in the best-performing models (bolded cells). LMS data have weaker predictive power but can oftentimes improve the performance of institutional data when combined. In contrast, survey data exhibits less utility; adding survey data to other data sources can even harm the prediction performance. One reason might be that pre-course survey did not accurately capture students’ skills as students tended to overestimate themselves with limited understanding of the course.

Figure 1 illustrates the fairness of different prediction results. Each square colors the quantified bias against a student subpopulation (FPR disparity or FNR disparity) under a specific combination of feature and target. Darker squares suggest larger biases and crossed out squares represent statistical significance ($\alpha=0.1$). Institutional data leads to higher FNR among disadvantaged student, suggesting that protected attributes (e.g., ethnicity) may induce identity-based bias. However, identity-free LMS data is also associated with such biases, albeit in the other direction (FPR). Moreover, students with low high school GPAs are no less harmed by unfair predictions than demographic minority groups.

These preliminary results suggest that collecting institutional data and, if possible, LMS data might be an overall cost-effective start for institutions to deploy student predictive analytics. To ensure fair outcomes, however, it is advisable to attend to cognitively or psychologically disadvantaged students

as much as to demographic minorities during the model building process. Future work will investigate how the biases illustrated in Figure 1 permeate through the predictive analytics pipeline.

Table 1: Prediction results for different student features and outcomes

Feature	Accuracy		False Positive Rate		False Negative Rate	
	Short	Long	Short	Long	Short	Long
Institutional	0.665	0.699	0.287	0.263	0.319	0.310
Click	0.637	0.626	0.357	0.340	0.263	0.343
Survey	0.567	0.542	0.344	0.447	0.277	0.330
Institutional+Click	0.699	0.714	0.287	0.279	0.275	0.286
Institutional+Survey	0.675	0.697	0.289	0.273	0.316	0.291
Click+Survey	0.625	0.623	0.367	0.371	0.295	0.338
Institutional+Click+Survey	0.706	0.713	0.297	0.274	0.271	0.272

Short: Course final grade above course median; Long: Next-year average GPA above course median

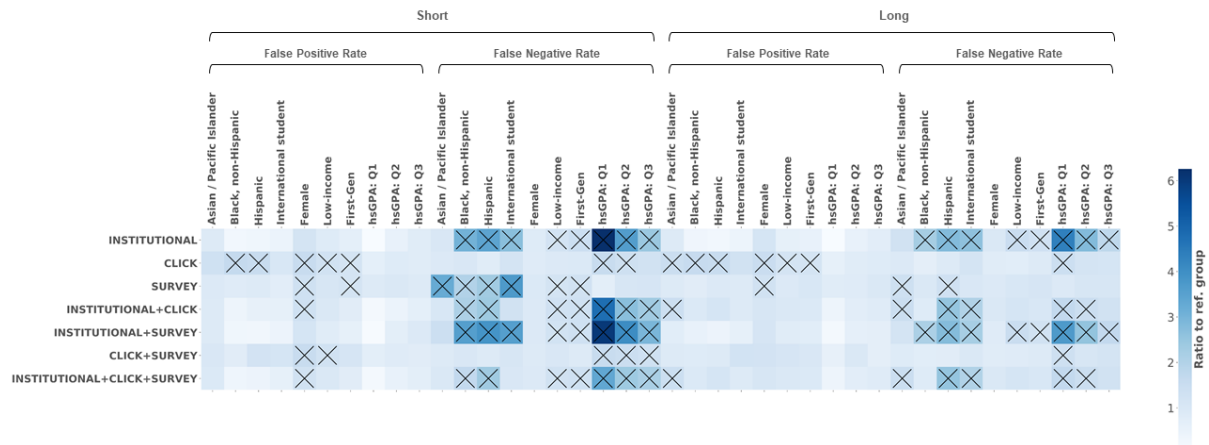


Figure 1: Bias against student subpopulations in different predictive models

REFERENCES

- Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *Internet and Higher Education*, 27, 1–13. <https://doi.org/10.1016/j.iheduc.2015.04.007>
- Hutt, S., Gardner, M., Duckworth, A. L., & D'Mello, S. K. (2019). Evaluating Fairness and Generalizability in Models Predicting On-Time Graduation from College Applications. In *The 12th International Conference on Educational Data Mining (EDM)* (pp. 79–88). Montréal, Canada.
- Ishitani, T. T. (2006). Studying Attrition and Degree Completion Behavior among First-Generation College Students in the United States. *The Journal of Higher Education*, 77(5), 861–885. <https://doi.org/10.1080/00221546.2006.11778947>
- Park, J., Yu, R., Rodriguez, F., Baker, R., Smyth, P., & Warschauer, M. (2018). Understanding Student Procrastination via Mixture Models. In *Proceedings of the 11th International Conference on Educational Data Mining (EDM)* (pp. 187–197). Buffalo, NY, United States. Retrieved from http://educationaldatamining.org/files/conferences/EDM2018/papers/EDM2018_paper_122.pdf
- Saleiro, P., Kuester, B., Hinkson, L., London, J., Stevens, A., Anisfeld, A., ... Ghani, R. (2018, November 13). Aequitas: A Bias and Fairness Audit Toolkit. Retrieved from <http://arxiv.org/abs/1811.05577>

Profiling students' mathematics motivation in 9th-grade: How does student motivation change over one academic semester?

Xiaoxue Zhang^{1 2}, James Middleton², Elizabeth Farley-Ripple¹, Zachary Collier¹, Amanda Jansen¹

¹University of Delaware, ²Arizona State University
veraz@udel.edu, jimbo@asu.edu, {enfr, collierz, jansen}@udel.edu

ABSTRACT: Student mathematics engagement and motivation are critical in secondary schools, particularly at the 9th-grade level. A validated survey of students' perception of learning environment and mathematics motivation was administered to 200 9th-grade students from 10 teachers' classrooms in Delaware and Arizona at the beginning and the end of the spring semester in 2018. In this paper, we examined how student motivation changes over one academic semester using cluster analysis and cluster transition analysis. We identified profiles of students' mathematics motivation. We also explored how students' mathematics motivation changed in response to the changes in the classroom learning environment over the academic semester in spring 2018. Results and further implications are discussed.

Keywords: Mathematics motivation; classroom learning environment; 9th-grade mathematics; cluster analysis; cluster transition analysis

1 BACKGROUND

The 9th-grade mathematics classes are often perceived as gatekeepers for students' academic advancement and future career opportunities (Graham & Chicas, 2015). Studying changes in students' mathematics motivation and what can contribute to the changes from the classroom learning environment in 9th grade is particularly important, given the significant influence mathematics motivation has on subsequent mathematics engagement and learning outcomes. The person-centered analytic approach, such as cluster analysis, can help complement the study of student engagement and motivation in attending to the diversity of student experiences identified through different profiles constructed by the multidimensionality of the construct simultaneously (Lawson & Lawson, 2013).

2 OBJECTIVES

The study has two goals: 1) examine profiles of students' mathematics motivation in 9th grade; 2) examine changes in students' mathematics motivation profiles over the semester, and in particular, how they respond to the changes in the classroom learning environment.

3 METHODS (PARTICIPANTS AND DATA COLLECTION MEASURES)

In Spring 2018, 200 9th-grade students from classrooms of ten teachers in six socio-economically and ethnically diverse schools in Delaware and Arizona participated in this study. This study used a validated survey measure on students' perception of the mathematics classroom environment, and mathematics motivation (Zhang, Middleton, Wiesel, Tarr, & Jansen, 2018) as student self-reflection survey is reliable in collecting student perceptions, and also convenient in gathering data from target participants. The survey adopts a 7-point Likert scale (where seven is rated as most positive and one as most negative) on all items. In January 2018, 200 students in participating classrooms filled out the surveys. In May 2018, the same 200 students were surveyed again.

4 DATA ANALYSIS AND RESULTS

4.1 Profiles of students' mathematics motivation

Four validated scales¹ consist of students' mathematics motivation: self-efficacy, interest, mastery goal, and performance goal. Cluster analysis was conducted based on these scales to examine the students' mathematics motivational profile. Five different cluster models ranging from 2 to 6 clusters were compared. Cluster validation methods, including elbow method, hierarchical dendrogram, k-means convergence criteria, and BIC fit statistics, nearly all suggested that the best model is the three-cluster solution for both time points.

Three student mathematics motivational profiles are identified (ranked from negative to positive): cluster 1: low motivation; cluster 2: average motivation with high performance goal; and cluster 3: high motivation with moderate performance goal (as shown in Figure 1 below). The identical student motivational profile found at these two different time points suggested the reliability of cluster results across time. The cluster patterns are also aligned with the mathematics motivation literature.



Figure 1: Students' motivational profiles at the beginning and end of the semester

4.2 Changes in student mathematics motivational profile in response to the changes in the classroom learning environment

Three validated scales² consist of students' perceptions of mathematics classroom learning environment: teacher support, peer support, and classroom performance goal structure. From paired sample t-test, a statistically significant difference ($p = .018$) was found for the change on peer support (an average increase of 0.17 on the 7-point Likert scale), but not teacher support (-0.01), or classroom performance goal structure (-0.07).

The cluster transition analysis on student motivational profile at the two time points revealed substantial changes among students' mathematics motivation over the semester. Among the 200 students, 69 (34.5%) moved to a different motivational profile, and more specifically, 40 (20%) students moved more negatively, and 29 (14.5%) students moved more positively (more details in

¹ The Cronbach's alpha for these four scales range from 0.80 to 0.91. Example survey items include 1) self-efficacy: I am confident that I can do an excellent job on math assignments; 2) interest: In general, when I work on math I have fun; 3) mastery goal: It's important to me that I thoroughly understand my math work, and 4) performance goal: It's important to me that I look smart compared to others in my math class.

² The Cronbach's alpha for the three scales range from 0.82 to 0.93. Example survey items include 1) teacher support: The feedback I have received from my teacher is valuable in this class; 2) peer support: My classmates in my math class care about how well I learn; and 3) classroom performance goal structure: My math teacher tells us how we compare to others.

Table 1 below). As indicated by the proportions of students who changed to a different profile, the change pattern was particularly salient among students who started with an average motivation.

Table 1: Students' motivational profile transition between the beginning and end of the semester.

Cluster	N or %	Post Cluster 3: High motivation with moderate performance	Post Cluster 2: Average motivation with high performance goal	Post Cluster 1: Low motivation	Grand Total
Pre Cluster 3: High motivation with moderate performance goal	n %	51 71%	12 17%	9 13%	72 100%
Pre Cluster 2: Average motivation with high performance goal	n %	16 21%	41 54%	19 25%	76 100%
Pre Cluster 1: Low motivation	n %	3 6%	10 19%	39 75%	52 100%
Grand Total	n	70	63	67	200

To investigate how students' motivation changed in response to the changes in the classroom learning environment, multinomial logistic regression was performed. The results suggested that students with negative motivational profile transition tended to blame on decreased peer support ($p = .002$) and greater emphasis on the performance goal in classrooms ($p = .001$), an indication of being left out or their contributions were not respected, while students with positive motivational profile transition tended to attribute to increased teacher support they received ($p < .001$).

5 CONCLUSION AND IMPLICATIONS

First, high volatility of mathematics motivation during one semester of 9th-grade mathematics classrooms has been found, particularly for students who began with a medium motivation. Second, teacher support and peer support may be influencing student mathematics motivation towards a certain direction more saliently. Finally, the person-centered analytic approach, such as the cluster analysis and its follow-up analyses, can be a useful way to explain student classroom experiences.

REFERENCES

- Graham, S., & Morales-Chicas, J. (2015). The Ethnic Context and Attitudes toward 9th Grade Math. *International Journal of Educational Psychology*, 4(1), 1-32.
- Lawson, M. A., & Lawson, H. A. (2013). New conceptual frameworks for student engagement research, policy, and practice. *Review of Educational Research*, 83(3), 432-479.
- Zhang, X., Middleton, J., Wiezel, A., Tarr, G., & Jansen, A. (2018). *Validation of an instrument for measuring math engagement, math motivation, and perception of the math classroom environment in secondary classrooms*. Tempe, AZ: Arizona State University.

Personality-Aware Educational Recommendations: Subjective vs Inferred Personality Traits

Yong Zheng and Archana Subramaniyan
Illinois Institute of Technology, USA
yong.zheng@iit.edu, asubramaniyan@hawk.iit.edu

ABSTRACT: Recommendation algorithms that were built based on the personality traits have been applied in the area of educations. The major challenge behind is the collection of personality traits which were usually collected from the questionnaires. However, the users may not be able to tell the truth in the surveys, which results in inaccurate personality traits and negative impact on the performance of recommendations. In this paper, we utilize a text conversation to infer personality traits, in addition to the subjective traits that were collected from user surveys. We examine the course project recommendations by using these subjective and inferred personality traits. Our experimental results can demonstrate the effectiveness of inferred personality traits in contrast to the subjective ones.

Keywords: personality trait; recommender system; virtual agent; text conversation

1 INTRODUCTION AND RELATED WORK

Recommender systems is a well-known for its capability of assisting user decision making by recommending a list of the items tailored to user preferences. Educational recommender systems have been utilized as one of the technology-enhanced learning methods, in order to suggest books for K-12 users [Pera, et, al., 2013], recommend after-school programs [Burke, et, al., 2012], or suggest appropriate citations [He, et, al., 2010] in paper writings. Personality-based recommendation models produce the item recommendations by taking advantage of the personality traits. These models have been demonstrated to be useful to alleviate the problem of new users [Rong, et, al., 2010], increase diversity in recommender systems [Chen, et, al., 2013] or generally improve the recommendation performance [Zheng, et, al., 2019]. User personality can be captured by the personality traits which can be represented by different personality frameworks. One of them is the big 5 framework [McCrae, et, al., 1992] which uses five dimensions (i.e., Openness, Conscientiousness, Agreeableness, Extraversion, Neuroticism). The most common way to collect the personality traits is user survey, e.g., the Ten-Item Personality Inventory (TIPI) [Gosling, et al., 2003]. However, users may not be able to tell us the truth through user surveys. For example, they could be shy or dishonest, or other reasons. It is better to infer the personality traits by using special models. In this paper, we examine the performance of course project recommendations by using both subjective and inferred personality traits, in order to figure out which way is better.

2 DATA COLLECTIONS

We collected our own data in the context of project recommendations [Zheng, 2018]. Each student was asked to fill a questionnaire by himself or herself. In the questionnaire, we present a total of 70

course projects to the students. Each student should select at least three liked and disliked ones, and provide an overall rating (in scale 1 to 5) to them. We have collected the data for two years. There is a total of 3,306 ratings given by 269 students on 70 items. In terms of the personality traits, we collected them in two ways. On one hand, we collected the **subjective personality traits**. More specifically, we develop the questionnaire based the TIPI to collect students' personality traits in the big five personality framework. However, the subjective personality traits may not be that reliable, e.g., students may be shy or dishonest to answer some questions, such as "I see myself as disorganized, careless". On the other hand, we used the text conversations to **infer personality traits**. This technique was introduced by an interview virtual agent [Li, et, al., 2017] which was developed by juji.ai. We collaborated with juji.ai and developed our own virtual agent. The agent was distributed to fresh-year students to learn their personal and academic preferences, such as their major, favorite classes, personal hobby, etc. As long as students can provide at least 1,000 words in textual chatting, the juji.ai system can automatically apply a computational model which uses rich linguistic cues (e.g., words, phrases, emoticons, and punctuations) to infer user traits. The system will deliver these traits in 35 dimensions, including Openness, Imagination, Adventurousness, Orderliness, Dutifulness, Extroversion, Friendliness, Impulsiveness, etc. In comparison with the big-five framework, we can obtain more dimensions by using the virtual agent. The major advantage of using the text conversations is that it provides a real-world and self-defined context (such as freshman at school, holidays, career, etc.) for conversations, and the personality traits can be inferred from the textual information without explicitly asking the questions related to any personal characteristics. The subjects will not feel embarrassed or offended during the process.

3 RESULTS, CONCLUSIONS AND FUTURE WORK

We use three mainstream personality-based recommendation models [Zheng and Subramaniyan, 2019] to evaluate the quality of the subjective and inferred personality traits. The 1st model is KNN based method [Hu, et, al., 2010] which uses personality traits to calculate the similarity of the users in the popular user-based collaborative filtering recommendation model. These similarities will be linearly combined with the similarities based on user ratings. The 2nd one is the matrix factorization model [Elahi, et, al., 2013] in which we extend the user-late factor vector by adding the vector representation of each personality trait. The last one is item splitting (iSplitting) [Baltrunas, et al., 2009] which was originally developed for context-aware recommendations. We can utilize the personality traits to split the items instead of using contexts. We evaluate these models by using the 5-fold cross validations, and present the recommendation performance in mean absolute error (MAE), and the normalized discounted cumulative gain (NDCG) in top-10 recommendations.

Table 1: Results of Recommendation Performance

	Algorithms	MAE	NDCG
Using Subjective Personality Traits	KNN	1.108	0.29
	MF	1.067	0.248
	iSplitting	1.067	0.255
Using Inferred Personality Traits	KNN	1.082	0.299*
	MF	1.054	0.256
	iSplitting	1.050*	0.264

The experimental results are shown in Table 1, while the values in bold present the best performed

MAE and NDCG results in the models by using subjective and inferred personality traits respectively. We use * to represent the significant and better results by using inferred personality traits based on 95% confidence level in comparison with the best models by using subjective personality traits.

We can observe that using the inferred personality traits can generally improve the recommendation performance in both MAE and NDCG in comparison with the algorithms using the subjective personality traits. More specifically, in the KNN based approach, the weight for personality-based similarities is 0.6 when we use the subjective personality traits. It is 0.7 when we use the inferred personality traits. The importance or the weight of the contributions based on the personality traits have been increased, if we use the inferred personality traits. It results in improved MAE and NDCG in the KNN-based models. Similar patterns can also be found in the MF-based and iSplitting based recommendation results. These experimental results reveal that the inferred personality traits are more reliable, and they can further improve the performance of personality-aware course project recommendations. We plan to examine the inferred personality traits in the task of group recommendations in our future work.

REFERENCES

- Baltrunas, L. and Ricci, F. (2009). Context-based splitting of item ratings in collaborative filtering. *Proceedings of the ACM Conference on Recommender Systems*, ACM.
- Burke, R. and Zheng, Y. and Riley, S. (2011). Experience Discovery: hybrid recommendation of student activities using social network data. *Proceedings of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems*. ACM.
- Chen, L. and Wu, W. and He, L. (2013). How personality influences users' needs for recommendation diversity? *In CHI'13 Extended Abstracts*. ACM.
- Elahi, M. and Braunhofer, M. and Ricci, F. and Tkalcic, M. (2013). Personality-based active learning for collaborative filtering recommender systems. *In Congress of the Italian Association for Artificial Intelligence*, pages 360–371. Springer.
- Gosling, S. D. and Rentfrow, P. J. and Swann Jr, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality*, Elsevier
- He, Q. and Pei, J. and Kifer, D. and Mitra, P. and Giles, L. (2010). Context-aware citation recommendation. *In Proceedings of International Conference on World Wide Web*, ACM.
- Li, J. and Zhou, M. X. and Yang, H. and Mark, G. (2017). Confiding in and listening to virtual agents: The effect of personality. *In ACM conference on Intelligent User Interfaces*, ACM.
- Hu, Rong and Pu, Pearl (2010). Using personality information in collaborative filtering for new users. *In Workshop on Recommender Systems and the Social Web*, ACM
- McCrae, R. R. and John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of personality*, 60(2):175–215, John Wiley & Sons, Inc.
- Pera, M. S. and Ng, Y.-K. (2013). What to read next?: making personalized book recommendations for k-12 users. *In ACM Conference on Recommender Systems*, ACM.
- Zheng, Y. and Subramaniyan, A. (2019). Personality-Aware Collaborative Learning: Models and Explanations, *The 33rd International Conference on Advanced Information Networking and Applications*, Springer.
- Zheng, Y. Personality-Aware Decision Making in Educational Learning (2018). *Proceedings of the 23rd ACM Conference on Intelligent User Interfaces*, ACM

Learning Analytics for Inclusive STEM Student Success

Xiaojing Duan, Alex Ambrose, Chaoli Wang, Kevin Abbott, Victoria Woodard, Kelley Young

University of Notre Dame

xduan@nd.edu, gambrose@nd.edu, chaoli.wang@nd.edu, kabbott@nd.edu,
vweber1@nd.edu, k.m.h.young@nd.edu

ABSTRACT: The challenge was to identify and help underserved and underprepared students in an introductory chemistry course to be retained and thrive in the college of science or engineering while supporting the general population. In this paper, we describe our methods for identifying these students, evaluating the impact of a special treatment program that was provided to a subset of those students, discuss our efforts to help the general population, and evaluate the short- and long-term impacts. In particular, we discuss a data-informed framework for analyzing student and outcome variables.

Keywords: STEM Retention, Learning Visualization Dashboard, Inclusive Pedagogy, Learning Analytics

1 INTRODUCTION & CONTEXT

It is often the case that students at a university find a specific course within an academic program challenging. These courses have been dubbed gateway courses because students feel they are being intentionally pulled from their desired program due to lack of success in these specific courses. The course of interest for this study is Introduction to Chemical Principles (Chem I). This course is the gateway course into all College of Science (CoS) and College of Engineering (CoE) programs and with about half of incoming first year students taking the course, it is the 2nd largest course on campus. This project began in the summer of 2018 and is ongoing. In the fall of 2018, 955 first year students enrolled in Chem I, with 45 students being admitted to the first Science & Engineering (S&E) Scholars program. This program was designed to close the achievement gap of underserved and underprepared students in the CoS. The project is currently evaluating the second cohort of the S&E Scholars program treatment impact on Chem I performance students. The intended audience for this work includes the dean and assistant dean of the CoS, the coordinator for the Chem I course and the instructors of the students in the S&E Scholars program. This project was staffed by university personnel from the Center for Teaching Excellence, Academic Technologies, Institutional Research, and the Department of Applied and Computational Mathematics and Statistics. Funding for the project was provided by an NSF Undergraduate Research Grant and Schlindwein Family International Research Grant.

2 CHALLENGES AND GOALS

At a medium-sized private religious institution, it is the goal to ensure all students stating an intention to study in a STEM field upon admission to the university, have the maximum opportunity to thrive in their desired field. More specifically, the university desires to help the underserved and underprepared special populations thrive while also maintaining the rigor demanded of students within the general population of the CoS and CoE. A second goal that the university established was

to determine if a treatment program aided students in thriving within the Chem I course, and helped them to thrive beyond the course. More specifically, we were tasked to determine if the treatment program had a long-term effect on retention within the CoS or CoE. However, because of the long-term nature of retention, and the desire to determine retention on a short-term basis, indicators of retention needed to be identified. These goals led us to our three research questions: (1) What historical data suggest Chem I and STEM majors are not inclusive at the University? (2) How can we close the achievement gaps and maximize all students' potential to thrive? (3) What are the short and long-term outcomes when comparing the treatment (S&E Scholars) and the control group (students with similar characteristics)?

3 DISCUSSION, CONCLUSIONS & FUTURE WORK

In conducting this work, we experienced three challenges in our implementation. First, we needed to redefine the usual binary STEM success to a multi-faceted measure of short term and long term outcomes. Second, we needed to get data permissions and pipelines that connected multi-source student data from admissions, registrar, and live LMS/Publisher HW data. Third, managing expectations and pacing culture change for faculty to utilize data driven decision making to act on real-time data and continuously improve course and exam design.

From the data that was collected we were able to create a data-informed methodology to help identify variables of interest for our evaluations in aiding students to thrive in STEM programs. These included, ask/test, then answer pre, during, and post CHEM I course evaluation questions in order to maximize inclusive STEM student thriving in the first year and semester gateway class. By combining data from multiple sources, an learning analytics platform was created to examine the inclusivity of Chem I and to identify significant attributes of students that struggle to thrive in the course.

In addition to creating a data driven measure for inclusivity, we also redefined success in a STEM major to include more than the binary outcome of retaining that major. Post Chem I indicators such as higher first semester GPAs and Chem II grades were indicators of a student thriving and were found to be good predictors of student retention in a STEM major in the long term.

Finally, we also created a performance visualization dashboard that allows deans and instructors to determine whether a boost (positive intervention) has an impact on a student's success. This dashboard includes options for the S&E scholars program treatment, a homework and exam item analysis and whether a student was boosted from the live grade book exam and homework tracking. This visualization tool could be used by others to identify students who are unlikely to thrive in a course, suggesting that a boost might be of help to their success.

In the future, our visualization will be utilized to identify students unlikely to thrive in a course and boosts will be applied. We will evaluate our ability to aid students to thrive in the classroom based on those results. Additionally, we can use similar methods to what we have described here to help students thrive in other STEM gateway courses usually taken in the first or second years (i.e. Calculus, Physics and Introduction to Engineering). Future work also includes plans to visualize and analyze item analysis question data to ensure valid and reliable homework and exams and identify and notify early student performance triggers that predict non-thriving outcomes.

Designing A Collaborative Problem-Solving Activity to Prepare Students for Flipped Classroom

Yuqian Chai, Xinyu Qi, Ling Li, Mansurbek Kushnazarov, Cheuk-Wang Yau,
Yifei Dong, Chi-Un Lei

The University of Hong Kong

yqchai@eee.hku.hk, andreaq@hku.hk, lingli1000@gmail.com, mansurbek@teli.hku.hk,
mcwyau@hku.hk, chloedong@teli.hku.hk, culei@hku.hk

ABSTRACT: Despite flipped classroom's potential positive influence on students' learning experiences, lack of student pre-class preparation has been a widely reported issue. To address this issue, we have designed a collaborative problem-solving activity to promote students' pre-class online engagement and to get them better prepared for flipped classroom learning. With this approach, students are motivated to engage actively with online learning materials, and instructor and/or tutors are enabled to provide timely support to students before class. Experiments have been conducted to verify the effectiveness of the activities, with results showing that student engagement in the experimental group was strikingly higher than in the control groups.

Keywords: Flipped Classroom, Collaborative Problem-Solving, Student Engagement.

1 INTRODUCTION

The flipped classroom (FC) approach has received remarkable consideration in higher education during the past few years. By moving didactic learning materials online and distributing them before class, FC allows face-to-face (F2F) class time for interactive learning activities. This instructional approach has been proven to positively influence students' learning performance. Nevertheless, students' poor preparation before class has been a major challenge in this area, and such inadequate preparation has a negative influence on students' in-class performance and overall learning outcomes. To resolve this issue, a collaborative problem-solving activity is proposed to better engage students in pre-class learning. This activity is student-centered, teacher-facilitated, and data-informed, with great potential to be generalized in flipped classroom learning.

2 ACTIVITY OVERVIEW

The collaborative problem-solving activity consists of 4 stages of implementation as depicted in Figure 1. The following elaborates on the teaching team's detailed execution in each stage:

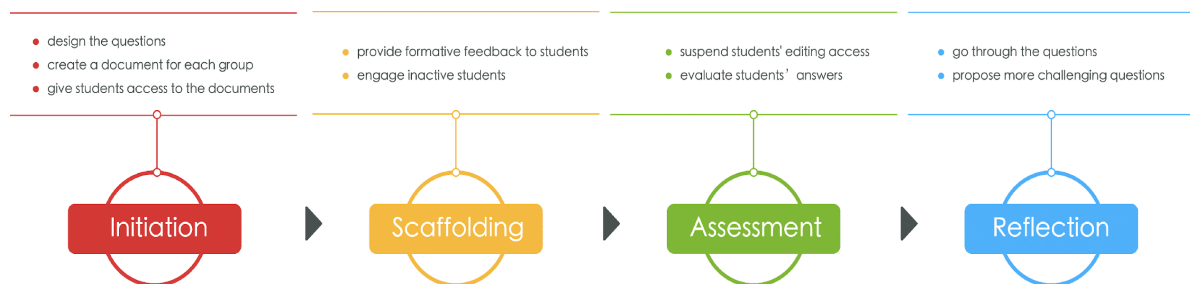


Figure 1: The workflow of the proposed collaborative problem-solving activity

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

- **Initiation:** designing high cognitive demand open-ended questions for students' collaborative problem-solving. The questions are distributed via an online collaborative editing tool (Google Docs in this study), which serves as a group learning space for students to discuss, comment on and construct responses.
- **Scaffolding:** providing support and feedback to student groups to facilitate their collaboration. Referencing each individual's editing records stored in Google Docs, instructor and/or tutors would perform two types of real-time interventions: 1)formative feedback to students' collective response to the open-ended questions; and 2)reminders to inactive students who are not engaged in group work.
- **Assessment:** assessing students' contribution, collaboration, and cognitive performance with reference to their responses and editing records.
- **Reflection:** resolving common issues among student groups to facilitate students' reflection, which can be followed by more challenging questions for knowledge consolidation.

3 EXPERIMENT AND RESULTS

We conducted experiments in an undergraduate General Education course that was conducted in a fully-flipped format with 123 students to verify the effectiveness of the proposed activity. In this course, students are required to watch a series of online lecture videos before attending the F2F class activities. Students from the third cohort is the experimental group, whom were assigned the collaborative problem-solving task before each F2F class. Students from the first two cohorts without such requirement serve as control groups. Their online engagement data was collected, and the video access rate data is displayed in Figure 2. In the heatmap, each row represents a student and each cell stands for the student's lecture video access rate (number of videos accessed divided by the total number of videos in the topic). The higher a student's video access rate is, the darker the corresponding cell is. As can be seen from the figure, students in the experimental group had higher online engagement level compared with the control groups.

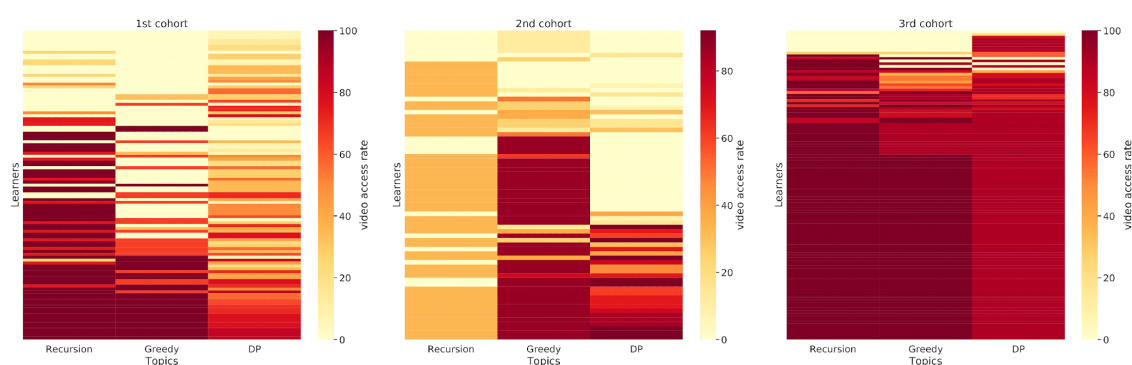


Figure 2: The heatmap of video access rate

REFERENCES

Akçayır, G., & Akçayır, M. (2018). The flipped classroom: A review of its advantages and challenges. *Computers & Education*, 126, 334-345.

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

Towards Better Grading: Promoting Grading Fairness with Assessment Decision Tree Visualization

Yuqian Chai, Xinyu Qi, Ling Li, Mansurbek Kushnazarov, Cheuk-Wang Yau,
Yifei Dong, Chi-Un Lei

The University of Hong Kong

yqchai@eee.hku.hk, andreaq@hku.hk, lingli1000@gmail.com, mansurbek@teli.hku.hk,
mcwyau@hku.hk, chloedong@teli.hku.hk, culei@hku.hk

ABSTRACT: Assessment fairness has been a challenging issue in educational practice, particularly for large-scale courses in which multiple graders are involved. In order to help guarantee grading fairness, this paper proposes an approach—‘assessment decision tree visualization’—that enables graders to identify error cases and adjust the grading if necessary. After implementing the approach into a university course with 136 students, results showed such visualizations could help teachers achieve fairer grading practices.

Keywords: Assessment Fairness, Decision Tree, Visualization

1 INTRODUCTION

Assessment fairness has remained challenging in educational research and practice for decades. Literature has suggested that grading transparency and use of rubrics could lead to fairer assessment. However, in a large sized class where multiple graders are involved, graders may have different interpretations of a same rubric, resulting in grading discrepancy. To ensure grading fairness, quality check is usually carried out, which is however a time-consuming and laborious task. In this study, we propose to use assessment decision trees based on students’ interaction data and grades to replace the manual check and achieve grading fairness without excessive effort.

2 EXPERIMENT

There is an ongoing trend of utilizing online interaction data in teaching and learning practices. It could offer an overview of students’ learning performances and habits without too much effort required. Subsequently, if we select students’ interaction data as the input and grades as the prediction target, we can build the assessment decision tree. The decision tree algorithm is a classical machine learning algorithm that can be applied in both classification and regression issues. One of the major advantages of the decision tree is that it is easy to interpret and visualize. Graders can adjust their grades by checking and comparing their assessment tree with others’.

2.1 Method

To better illustrate the process, we have conducted an experiment in an undergraduate level course with an enrollment size of 136. In the course, students needed to complete a graded activity which required them to work in small groups and answer open-ended questions in an online shared document. The submissions are then graded by ten graders based on each individual student’s

contributions to the collective response. In other words, there should be a high correlation between students' interactions and grades. During the activity, students can either edit the documents directly to answer questions or create comments to discuss with teammates. Both actions are considered as contributions. Subsequently, we collected number of participated questions (*#questions*), number of comments (*#comments*), and total number of contributed words (*total*) as input features. Since there are four grade scales for this activity (0, 5, 7, 10), the decision tree we built is a classification tree. The average sample size of each grader's assessment decision tree is 68.

2.2 Results

We built each grader's assessment decision tree and two of them are presented in Figure 1. It is clear that Grader 2 was stricter than Grader 1 as students on the right tree are less likely to receive the highest score indicated by the root node. Specifically, Grader 1's decision tree indicates multiple scenarios in which students could receive the highest score (grade_10), while Grader 2's decision tree only suggests one of such scenarios with a small sample size of 3. The decision tree visualization was presented to the graders and most of them have modified their grades accordingly.

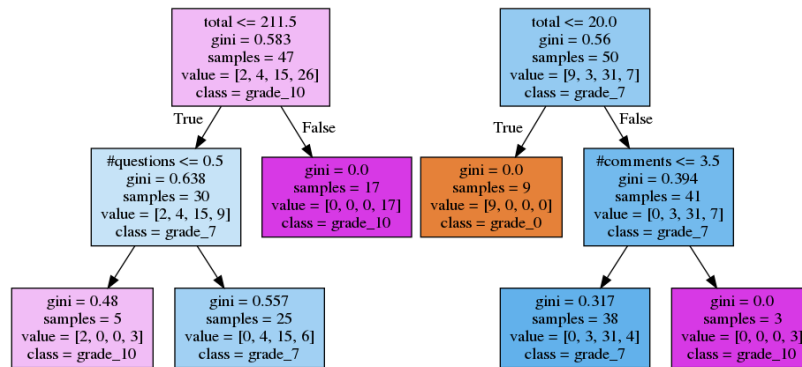


Figure 1: The assessment decision trees of Grader 1 (left) and Grader 2 (right)

3 LIMITATIONS

Despite the advantages of assessment decision trees, there are several limitations while applying this method. First of all, it is particularly applied in the evaluation of text-input participation. For other types of tasks, students' interaction data may not have a tied relationship with grades. In these cases, teachers can manually label some cognitive features, for example, frequency of grammar mistakes (seldom, sometimes, frequently). By doing so, the assessment decision tree can help verify the fairness in grading. Additionally, there is a sample size requirement for the decision trees, but it would not be an issue for large-scale courses.

REFERENCES

- Tierney, R. D. (2013). Fairness in classroom assessment. *SAGE handbook of research on classroom assessment*, 125-144.
- Magerman, D. M. (1995, June). Statistical decision-tree models for parsing. In *Proceedings of the 33rd annual meeting on Association for Computational Linguistics* (pp. 276-283). Association for Computational Linguistics.

Fertile breeding ground for learning analytics at scale: the KU Leuven approach

Tinne De Laet, Tom Broos, Inge Wullaert, Anneleen Cosemans, Katleen Craenen
KU Leuven
{tinne.delaet, tom.broos, inge.wullaert, anneleen.cosemans,
katleen.craenen}@kuleuven.be

ABSTRACT: Realizing learning analytics at an institutional scale is challenging. This poster presents the exemplary approach of KU Leuven, who stimulated by a set of learning dashboards created bottom-up, elaborated a policy around educational technology and installed a strategic plan to create a fertile breeding for learning analytics initiatives, scaling promising initiatives up to an institutional scale, and anchoring them in institutional processes and practices.

Keywords: learning dashboards, scalability, institutional change, policy

1 BOTTOM-UP LEARNING DASHBOARDS

Within two European projects [STELA](#) and [ABLE](#), four bottom-up learning dashboards were created that aimed at supporting the interaction between student advisor and students and self-reflection of students (Figure 1). LISSA [Charleer et al. 2018] was designed to support the conversation between student advisor and students based on an intensive user-centered design project. To support the self-reflection of students, three self-serving dashboards were designed and deployed: [LASSI](#) around learning skills [Broos et al. 2019], [REX](#) around academic achievement [Broos et al. 2019], and [POS](#) for aspiring students [Broos et al. 2019]). These dashboards were the result of a strong collaboration between a multidisciplinary team of researchers and practitioners. At the end of the project the dashboards were piloted in 26 programs within KU Leuven, reaching more than 4000 students and 120 student advisors.

2 POLICY AND STRATEGIC PLAN TO SUPPORT EDUCATIONAL TECHNOLOGY

While the learning dashboards were piloted at a large scale, strongly supported by student advisors and students, and well-received by KU Leuven policy makers, the project struggled for continuation and embedding in actual university practices and processes. At the same time, and inspired by the experience of the learning dashboards and other innovative projects struggling to scale to an institutional level, KU Leuven policy makers elaborated a policy around educational technology name [GoingDigital](#). This policy aimed at using educational technology such that it facilitates collaborative learning and multi-campus education and broadens the international reach. The policy plan also named 10 short term goals, including the scaling up of the developed learning dashboards.

The policy plan was later on translated to a strategic plan that provided concrete stimuli to realize educational technology, including project financing. The project financing is built around three phases, inspired by well-known maturity models: 1) the innovation phase to stimulate innovative

bottom-up initiatives, 2) the scaling-up phase, relying on a strong collaboration with institutional services (IT, educational policy, student services, educational support services, ...) to analyse if and how the best bottom-up initiatives can be scaled up, and 3) the actual upscaling and anchoring of the initiatives.

The poster will share the first experiences with the strategic plan, focused on the learning dashboards, which are currently in the scaling-up project phase. The poster will discuss both the engagement of diverse stakeholders, embedding in IT infrastructure university-wide, and policy making.

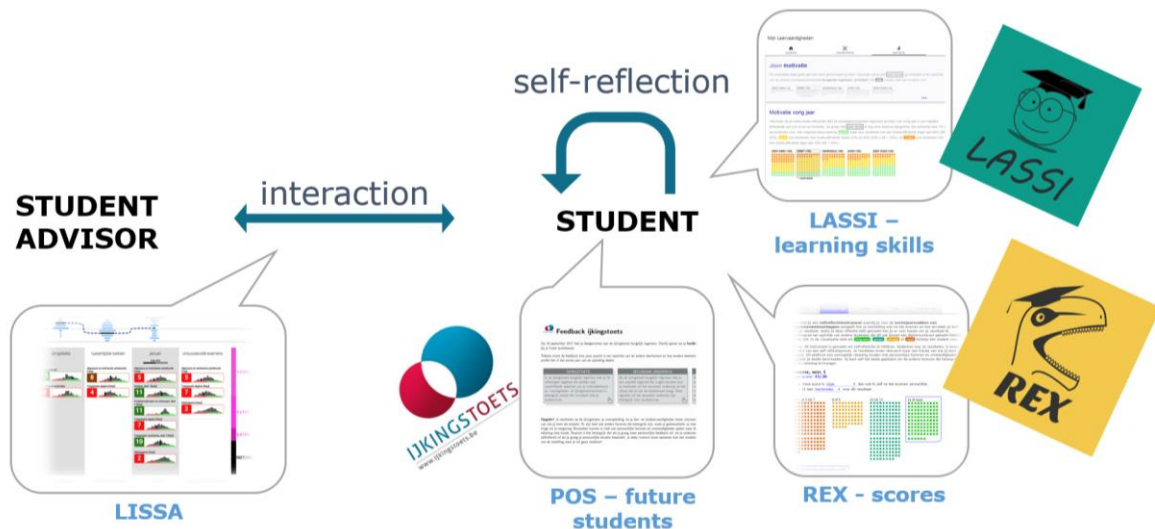


Figure 1: Overview of the learning dashboards at KU Leuven developed within two European projects: [STELA](#) and [ABLE](#). The dashboards support the interaction between student advisor and student and self-reflection. For demo's check [LASSI](#), [REX](#), and [POS](#)

REFERENCES

- Broos ,T., Pinxten, M., Delporte, M., Verbert, K., & De Laet, T. (2019) Learning dashboards at scale: early warning and overall first year experience, *Assessment & Evaluation in Higher Education*, DOI: 10.1080/02602938.2019.1689546
- Broos, T.; Verbert, K.; Langie, G.; Van Soom, C.; De Laet, T. (2018) Multi-institutional positioning test feedback dashboard for aspiring students: lessons learnt from a case study in Flanders; ACM; LAK, Date: 2018/03/05 - 2018/03/09, Location: Sydney, Australia; Proceedings of the Eighth International Learning Analytics & Knowledge Conference; 2018; pp. 1 - 6
- Charleer, S., Moere, A. V., Klerkx, J., Verbert, K., and De Laet, T. (2018). Learning Analytics Dashboards to Support Adviser-Student Dialogue, *IEEE Transactions on Learning Technologies*, 11(3), 389-399, <http://dx.doi.org/10.1109/TLT.2017.2720670>

Towards Instructor-based Predictive Learning Analytics with LAGradebook

Jesse Eickholt, Chris Phillips
Department of Computer Science
Central Michigan University
eickh1jl@cmich.edu

ABSTRACT: This practitioner poster presents a means to perform instructor-/course-level predictive learning analytics. Predictive learning analytics aims to predict student outcomes and enables instructors to target their interventions to those most in need. The cost and institutional support required by many existing systems for predictive learning analytics may place their adoption outside the reach of many instructors or institutions. To combat this limitation, we have developed an extension to LAGradebook. This extension allows instructors with little modeling knowledge to create and apply predictive models to their courses without the cost or requirement of large, institutional systems.

Keywords: predictive learning analytics, instructor-level, course outcome modeling

1 INTRODUCTION

Predictive learning analytics aims to predict student outcomes (e.g., final course grade) and can provide early warnings that identify students at risk (Barber & Sharkey, 2012; Siemens, Dawson, & Lynch, 2013). Such information, if provided in time, allows instructors to intervene on the student's behalf and potentially lead to better student outcomes. Some have argued that "analytic and predictive models need to be reliable and valid at the scale beyond the individual course or cohort" (Ferguson *et al.*, 2014). Arguably, there are advantages to larger, more generalizable learning analytics tools but cost or institutional adoption can be a limiting factor. To provide wide access to predictive learning analytics, options need to exist for enterprising instructors that lack institutional data or support.

2 FRAMEWORK AND PROPOSED WORKFLOW

To increase instructor level access to learning analytics, we have created LAGradebook. This is a standalone application that ingests raw, student assessment data in the form of a tab separated file (e.g., a gradebook dump from a learning management system) and creates a rich Excel spreadsheet. The spreadsheet supports a number of comparative analytics (e.g., relative overall performance or by assessment item, top performance counts) but the application previously lacked the capacity to directly support predictive analytics. To provide this support, an additional worksheet is now generated by LAGradebook that presents the user with the option of selecting and labeling assessment items and specifying a target value (e.g., a letter grade, final percentage, completion status). The selected assessment items are normalized as z-scores and packaged with the target values to produce training data that can then be submitted to a web service for model generation. Figure 1 shows the feature generation worksheet and the generated feature data.

The screenshot shows an Excel spreadsheet with the following data:

Model Name	Type	Regression ID	Inc. in Model	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
CS 101, Regression	Regression	9.584999,1.132127192	9.584999													
9.464999,1.024671052		9.464999														
7.795999,0.986293859		7.795999														
9.559999,1.016995614		9.559999														
6.854999,1.016995614		6.854999														
8.432999,0.890350877		8.432999														
8.700999,0.932565785		8.700999														

Instructions: Add a model name, select model type, specify targets and choose assessment items to include in the model. Then copy the contents of the cell to the right and paste in at

Figure 1: Feature generation worksheet of LAGradebook with training data highlighted

The webservice makes use of Scikit-Learn¹ to build a small multilayer perceptron and returns the weights that parameterize the model along with metadata that specifies which assessment items were used to build the model. In later iterations of the course, the instructor can map the previous assessment items to entries the gradebook. Predictions can then be made by providing the model. The predictions appear as an additional worksheet in the generated LAGradebook.

3 CONCLUSION

Early identification of at-risk students enables instructors to target their interventions to those most in need. Several large learning analytics systems have been developed that successfully use institutional data and operate on the institutional level. Given the cost and institutional support required by these tools, they may be outside the reach of many instructors. To overcome this limitation, we have developed an extension to an existing tool that supports learning analytics on an instructor-/course-level. The extension allows instructors easily leverage previous assessment data in a course to predict future student outcomes.

REFERENCES

- Barber, R., & Sharkey, M. (2012, April). *Course correction: Using analytics to predict course success*. Paper presented at the 2nd International conference on Learning Analytics and Knowledge. <https://doi.org/10.1145/2330601.2330664>.
- Ferguson, R., Clow, D., Macfadyen, L., Essa, A., Dawson, S., & Alexander, S. (2014, March). *Setting learning analytics in context: Overcoming the barriers to large-scale adoption*. Paper presented at the 4th International Conference on Learning Analytics and Knowledge. <https://doi.org/10.1145/2567574.2567592>
- Siemens, G., Dawson, S., & Lynch, G. (2013). Improving the quality and productivity of the higher education sector. *Policy and Strategy for Systems-Level Deployment of Learning Analytics*. Canberra, Australia: Society for Learning Analytics Research for the Australian Office for Learning and Teaching.

Getting staff ready for learning analytics: preliminary findings from the OfLA Erasmus+ project

Ed Foster

Nottingham Trent University
Ed.foster@ntu.ac.uk

Peter Crowson

Nottingham Trent University
Peter.crowson@ntu.ac.uk

Pieterjan Bonne

Artevelde University College
Pieterjan.bonne@arteveldehs.be

Rianne Bouwmeester

University Medical College Utrecht
R.A.M.Bouwmeester@umcutrecht.nl

ABSTRACT: Researchers from the OfLA Erasmus+ Project (2018-2021) conducted a series of staff interviews with tutors, advisers and academic managers to understand how they used data, including data from learning analytics, as part of student success interventions. The researchers found that learning analytics data was being used to augment the advising process, but that further work was needed to better integrate it into staff practice.

Keywords: institutional adoption, student success, staff interventions, early warning systems, personal tutors, academic advisers

1 INTRODUCTION

We would argue that the major challenge for learning analytics is effective institutional adoption. This may be particularly the case where learning analytics is used to support student success and reduce incidences of early ‘drop out’ where a wide range of staff may be involved in supporting students. Onwards from Learning Analytics (OfLA) is an Erasmus+ funded research collaboration between three partners (Nottingham Trent University (NTU), Arteveldehogeschool (AHS), University Medical Center Utrecht (UMCU). It builds on previous work conducted in the ABLE & STELA Erasmus+ projects. The project is investigating the role that staff play, particularly personal tutors/study advisers, in student success. The project is looking at student support processes in three stages: **trigger/prompt, communication and intervention.**

2 METHODOLOGY

During the 2018-19 academic year, the research team devised a common interview script and conducted interviews with staff in relevant roles to understand existing approaches to supporting students at risk. This work was conducted to inform pilots in the 2019-20 year and capture examples of best practice. The interviews were conducted face-to-face or via telephone/videoconferencing. Eighteen staff members at Arteveldehogeschool (AHS) were interviewed in a range of roles including study coaches, managers, policy workers and a counsellor (May-June 2019). At NTU, twelve personal tutors were interviewed (April-May 2019). They were selected as recipients of student ‘no-engagement’ alerts generated by NTU’s learning analytics platform (the NTU Student Dashboard).

3 FINDINGS

3.1 Trigger/ Prompt

Staff in AHS used a range of strategies to try and identify students at risk of early departure: these included background, engagement with studies, or an observed absence of belonging (through staff observation, there is no learning analytics system currently in place). Tutors at NTU reported that they generally found it useful to receive a 'no-engagement' alert as it helped them consolidate their general perceptions. *"It gives us some data, some numbers to it, rather than just being a hunch or something like that, so I think it's good to know."*

3.2 Communication

AHS staff reported that their preferred communication method was to use email and invite students to attend a one-to-one conversation. In addition, some rang students, particularly when they perceived that the student needed contacting urgently. Staff also actively sought students out during or after classes. One of the most challenging issues was the question of escalation when a student failed to respond to initial communication. Several interviewees reported that it was hard to know whether to leave students alone, or use more urgent forms of communication. At NTU, staff described benefits from the no-engagement alerts. For example, tutors were able to distance themselves from the alert *"I think it's useful to us because it's a prompt for everybody. We just say, almost apologetically to the students, this is flagged up. It's not anything about you. We're required to offer you support."* Whilst this approach may be a little disingenuous, it potentially reduces some of the stresses associated with the tutor/student power relationship.

3.3 Intervention

AHS staff were generally confident that there were well-articulated routes to further specialist support. However, there were problems with the existing model, for example concerns about resourcing and a lack of evidence-based guidelines for study guidance *"Every coach works with the best intentions, but do we reach our goals? If not, how can we adjust?"*. NTU staff reported using the NTU Student Dashboard in their tutoring direct with their students, reporting that it's *"a good start to the conversation"*, *"I think the use of Dashboard at the one-to-ones with students, reviewing their engagement, is really useful because it can give them a bit of a reality check and it shows that we're actually looking at how engaged they are."*

4 CONCLUSION & FUTURE FOCUS

This preliminary piece of work shows that there is scope for integrating data from learning analytics into student success practices. However, significant barriers remain. Staff have a strong sense of agency about their own skill at identifying students at risk through observations and interactions. The data from learning analytics did not always align with their perceptions and so they can be reluctant to integrate it into their schema of work. Further staff development is needed in existing areas such as coaching and communication skills, but also new specialist fields such as data literacy. In 2019/20 the researchers are expanding the work to experiment with specific interventions, testing both the efficacy of the approach and student/staff perceptions of the trials.

Student-centered Development of Learning Analytics at an Higher Education Institution

Jiri Lallimo
Aalto University
jiri.lallimo@aalto.fi

Amanda Sjöblom
Aalto University
amanda.sjoblom@aalto.fi

ABSTRACT: The emerging role of learning analytics tools for a university setting has been a key point of focus in a wider-scope project for supporting studying and learning at a Finnish Higher Education institution. The project has taken a user-centric co-design approach, and involved a variety of service design methodologies to investigate user needs. This has been followed by developing and piloting solutions in co-operation with multiple stakeholders: students, teachers, and many university services. The goal has been to discover ways to implement learning analytics tools to support students in a higher education setting, which requires considerable independent effort from the students in managing their own study paths. Here we aim to describe some of the key methodologies of co-design and the findings thus far, as well as outlining some of the primary directions and challenges for the future. The key implications highlight the benefits of a user-centric approach to designing effective tools for supporting both individual and organisational level of learning and teaching.

Keywords: service design; learning analytics, learning dashboards; study path; co-design

1 INTRODUCTION AND CONTEXT

In Finnish universities, students typically have rather a lot of freedom in constructing personal study paths compared to universities or programmes that appoint ready-made schedules. This freedom is provided to motivate students to find their own meaningful paths. However, this may also impose additional challenges as a high level of planning and managing of study contents, work load and schedules is required from the students. The aim of this study is to investigate the co-design and the applicability of learning analytics (LA) tools to support studying, and how these tools can aid students in managing and understanding their studies and learning. Given that student success and retention are related to study habits (e.g., Robbins, Lauver, Davis, Langley, & Carlström, 2004) and various aspects of study experience (Lizzio, Wilson, & Simons, 2002), we aimed to find ways to support organised study habits and study-related self-efficacy (Hailikari, Tuononen, & Parpala, 2018). We investigated the users' view of the support they would find beneficial for their studies, for what they would want to utilise analytics, and in what format.

2 METHODOLOGY

A service design approach applying a Double Diamond design model (British Design Council) was used to frame the investigation of user needs, and to develop the concept for LA supported student

dashboards. All relevant stakeholders (students, teachers, services) were involved throughout the process. In the Discovery phase, numerous interviews, questionnaires, meetings and workshops were organised to investigate the stakeholders' views and needs. This resulted in an outline of the elements of the ideal study path and the potential risk points, highlighting the parts where further support is needed. In the Define phase, by using methods such as fictional personas, customer service journey – maps, interviews and workshops, the goal was to formulate the LA student dashboard concept and its features, and to co-design and interact with prototypes. With students, we focused on how different tools could address various problem points, addressed data privacy concerns and other ethical issues, and investigated in what formats students would use the tools. During the Development phase, the prototype of the dashboard was developed based on the findings from the previous phases. In the current Delivery phase, the prototype is being piloted and developed for a production version.

3 RESULTS AND CONCLUSIONS

Stakeholders have been involved at each stage of the service design process, leading to an improving LA concept for supporting students. The iterative development process, in which key stakeholders' needs can be examined and answered at multiple stages, takes us towards a user-focused LA co-design, ensuring that user needs and LA capacities meet. Our findings indicate that the key challenges for students are the formation of an effective study plan that supports well-aligned learning and workload management. Students found calendar planning, LMS activity planning and monitoring, and course suggestion tools, which used LA for mapping study plans, interests, and courses, particularly useful. Comparison of own progress or success to other students' was found unhelpful. The concepts for the tools, their functions, presentation and scope were developed at each stage together with the stakeholders, resulting in a design increasingly well matched to user needs, leading to better approval from the user together with increasing likelihood to be used. It is important to notice that in this paper the focus has been on the co-design between users and main developers of LA. In the background, the development of student dashboard has required a dense collaboration between LA developers and pedagogical, IT, data and juridical expertise. The process of developing LA and the student dashboard proceeds with the current piloting phase. The process described here has demonstrated that with the users in a central focus, LA can be implemented to create added value to the university by including the relevant stakeholders' needs and perspectives and by motivating them to use new tools and practices.

REFERENCES

- Design Council. The Design Process: What is the Double Diamond? [Online]. Retrieved Nov 1st 2019. <https://www.designcouncil.org.uk/news-opinion/design-process-what-double-diamond>
- Hailikari, T., Tuononen, T., & Parpala, A. (2018). Students' experiences of the factors affecting their study progress: Differences in study profiles. *Journal of Further and Higher Education*, 42(1), 1-12.
- Lizzio, A., Wilson, K., & Simons, R. (2002). University students' perceptions of the learning environment and academic outcomes: implications for theory and practice. *Studies in Higher Education*, 27(1), 27-52.
- Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130(2), 261-288.

Towards a modular and flexible Learning Analytics framework

Yves Noël^{1,2}, François Bouchet¹, Roland Mergoil², and Vanda Luengo^{1,2}

Sorbonne Université - CNRS UMR 7606, LIP6 Paris France¹,

Sorbonne Université – CAPSULE, Paris France²

yves.noel@sorbonne-universite.fr

ABSTRACT: This paper introduces a Learning Analytics platform which aims at being modular, evolving and flexible. The general framework architecture is completely independent from the digital systems to which it is connected. It collects learning data of various origins in data storages. Then it extracts a subset of the data which is aggregated into a data warehouse. Finally, these data are processed through various algorithms. Such a framework reinforces the control of data integrity in an experimental context and allows the students to refine the authorizations they give about their data. These data processing lead to indicators that will be used in student and teacher dashboards allowing a clear and fast access to learning information. In a second step, the platform will compute student profiles, facilitating the design of adaptive courses for each student.

Keywords: Modular framework, Trusted learning analytics, Student consent

1 INTRODUCTION

Public French Universities welcome students with increasingly diverse profiles and face large population. To cope with this, Sorbonne University wants to offer to students adapted learning paths, give teachers a tool to adapt learning activities to the student needs and inform the students about their learning and commitment. For this purpose, we propose to develop a whole coherent framework centered on Learning Analytics. Such a platform has to be evolving and resilient because the digital environment is changing quickly. The pedagogical objectives can be achieved only if the acceptability by the stakeholders is strong. Furthermore, and in order to allow the appropriation of the framework it is built following ethical rules co-design with the main stakeholders.

2 THE FRAMEWORK

The platform framework consists of several interconnected components (see Figure 1). Since digital learning traces can have very different origins and formats, we have planned a first modular layer to collect and store these heterogeneous traces. In practice, this layer consists of a set of Learning Record Storage (LRS) databases, each LRS being used for storing the learning traces (using xAPI standard) of a given source. At first, only two LRS will be created: the one storing the learning traces from Moodle and the one storing the traces from the video platform of our university. We may eventually add a 3rd LRS associated with a serious game platform.

These LRS data will be translated (to Caliper¹ standard) to be associated in a Learning Record Warehouse (LRW) data warehouse ensuring the integrity and consistency of stored traces. Some of the traces collected are very detailed and are not intended to be associated in the LRW; they are only stored in the LRS. Although these data are useful for the source that generated them, they could be only unnecessary noise due to the instability or reliability of the source that can be an experimental source. The LRW also collects data from Student Information Systems (it is not possible in an LRS). Finally, a Learning Analytics Processor² (LAP) will process data from the LRW to compute indicators used in dashboards, or models used by the LAP itself.

The implementation of the platform does not only mean the technical deployment of the chosen architecture, but also the construction of a framework to ensure that the digital traces collected and the algorithms, comply with national and European regulations on the protection of data (GDPR). The platform framework modularity allows students to refine the authorizations they give about their data: the authorization can be specific to a source (Moodle, video platform, ...), and also restricted to a purpose (collect only, analysis for the student or analysis for the teacher/university). This fine control exercised by the student should increase the adoption of our platform by the learners. In addition, we have planned to provide ethical rules to our system. These rules will be co-constructed with the different actors involved (in particular students). We aim to implement trusted Learning Analytics following the DELICATE checklist as described by (Drachsler and Greller 2016).

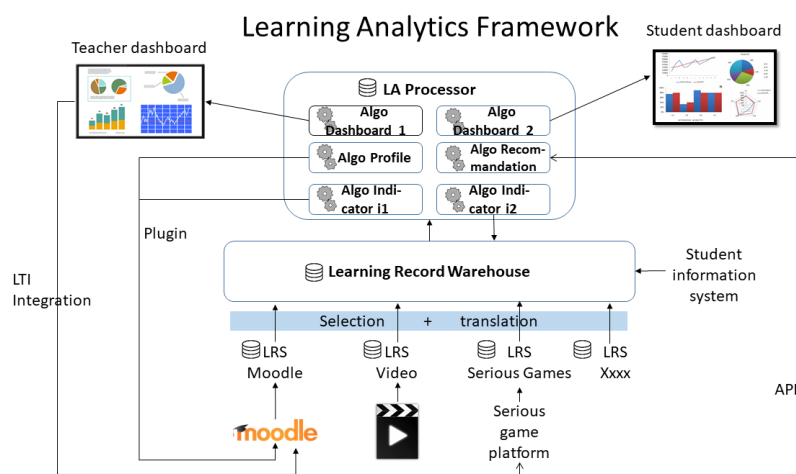


Figure 1: Schematic view of the Learning analytics framework architecture

REFERENCES

- Drachsler, H. & Greller, W. (2016). *Privacy and Analytics – it's a DELICATE issue. A Checklist to establish trusted Learning Analytics*. 6th Learning Analytics and Knowledge Conference 2016, April 25-29, 2016, pp. 89-98. Edinburgh, UK.
DOI: <http://dx.doi.org/10.1145/2883851.2883893>.

¹ <https://www.imsglobal.org/activity/caliper>

² <http://apereo-learning-analytics-initiative.github.io/LearningAnalyticsProcessor/>

Development of a Data-Oriented e-Learning Platform Based on the Community of Inquiry Framework

Xinyu Qi, Yuqian Chai, Ling Li, Mansurbek Kushnazarov, Cheuk-Wang Yau,
Yifei Dong, Chi-Un Lei

The University of Hong Kong

andreaq@hku.hk, yqchai@eee.hku.hk, lingli1000@gmail.com, mansurbek@teli.hku.hk,
mcwyau@hku.hk, chloedong@teli.hku.hk, culei@hku.hk

ABSTRACT: Community of inquiry framework has received increasing popularity in e-learning related studies in the past decade. In this prototype report, we present a data-oriented e-learning platform based on the community of inquiry framework. The main functions include student grouping, assignment creation, activity digest and student notifications, which are all essential steps to build and maintain an online community of inquiry. By utilizing students' background information and their learning behavioral data, the platform generates insights to inform teachers on the effectiveness of learning design and implementation.

Keywords: Community of inquiry, collaborative problem-solving, learning communities

1 INTRODUCTION

Effective learning experiences require learners to actively and collaboratively engage in exploring, creating meaning and confirming understanding. Such requirements are best represented in the community of inquiry framework. As depicted in Figure 1, this framework consists of three essential elements: social, cognitive, and teaching presence. It is at the intersection of all the three presences where students achieve meaningful educational experiences.

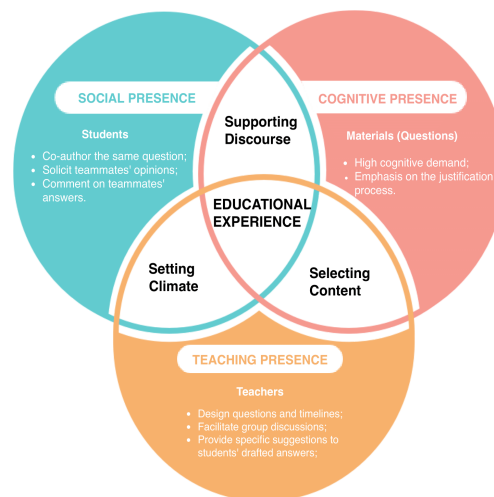


Figure 1: The community of inquiry framework

Based on the three requirements, several detailed specifications can be elicited and should be satisfied to ensure the online communities of inquiry are effective. Specifically, social presence is achieved through collaborative problem-solving and discussions; cognitive presence depends on the design of high cognitive demand tasks; and teaching presence can be realised through clear instructions, in-time support, and individual feedback. Considering these specific requirements, we designed a data-oriented platform that integrates all requirements into the activity design process.

2 PLATFORM DESIGN

The platform is designed to construct online communities of inquiry with collaborative editing documents (e.g., Google Docs). Compared to online discussion forums which are commonly used for online community building, collaborative documents offer more editing features, and could enhance group cohesion. The activity is group-based rather than in a whole-class setting, so each member has the opportunity and responsibility to contribute. The following elaborates on the main steps of designing activities using the platform.

- **Grouping:** As the foundation of social presence, grouping of students could be a daunting task when the class size is large and student body is diverse. The platform offers several parameters to enable teachers to group students with different strategies: maximizing or minimizing diversity, grouping based on their gender, major, year of study, etc.
- **Assignment Creation:** To satisfy the cognitive presence requirement, it is recommended to design high cognitive demand open-ended questions to facilitate the building of communities of inquiry. Using the assignment creation function shown in Figure 2(a), teachers only need to input the questions, and the platform will automatically generate collaborative editing documents for each student group and distribute the documents to their members.
- **Activity Digest:** Teaching presence requires in-time support and feedback, which can be very challenging to implement in large sized classes. The activity digest function offers a summary table of each group's progress and interaction level as shown in Figure 2(b). With such report, teachers can visualize the groups' performances, and over time, observe students' learning patterns and provide targeted intervention.
- **Student Notifications:** With the activity digest report, teachers can target groups or individuals to intervene with, and provide in-time support and feedback. Specifically, by allowing teachers to customize the parameter of idle students/ groups, as demonstrated in Figure 2(c), the platform can send reminders and follow up emails with personalized messages automatically to bring inactive students back on track.

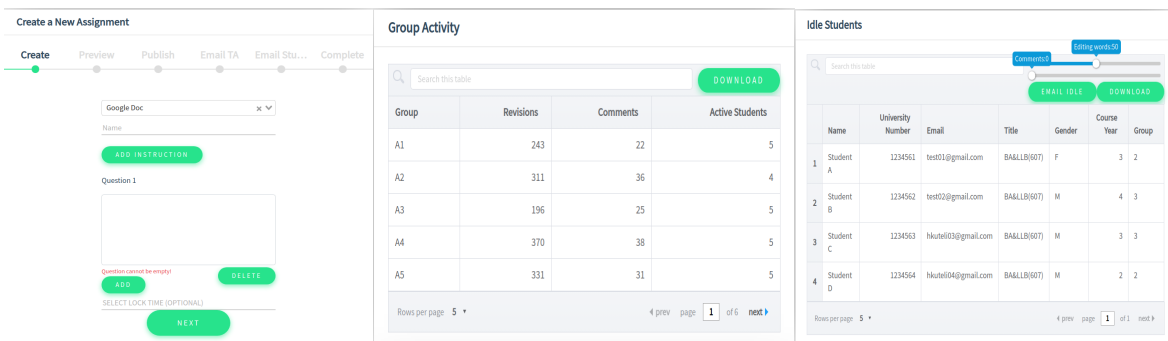


Figure 2: Screenshots of the platform

(a) Assignment Creation (b) Activity Digest (c) Student Notifications

REFERENCES

- Arbaugh, J. B., Cleveland-Innes, M., Diaz, S. R., Garrison, D. R., Ice, P., Richardson, J. C., & Swan, K. P. (2008). Developing a community of inquiry instrument: Testing a measure of the community of inquiry framework using a multi-institutional sample. *The Internet and Higher Education*, 11(3-4), 133-136.

Supporting online learners' awareness through weekly study reports for self-regulation of learning

Min Qu

Udacity

min.qu@udacity.com

Héctor J. Pijeira-Díaz

University of Oulu

Hector.PijeiraDiaz@oulu.fi

Yutian Ma

The University of Hong Kong

yutian95@connect.hku.hk

ABSTRACT:

Monitoring and being aware of learning processes continuously during a whole learning journey can be especially challenging for online learners, who usually have a busy life. However, online learners' trace data can be seamlessly recorded and used to support awareness. To that end, this poster introduces a weekly study report currently under development for the Udacity platform, which ultimately aims to support learners' self-regulation towards their online learning. The future of work is also discussed.

Keywords: learning analytics, study report, awareness, self-regulated learning, monitoring, trace data, data visualization

1 PROBLEM

During the learning journey of an online course student, it is important to support their awareness for them to better monitor their learning progress. Monitoring is central to self-regulated learning (SRL), which is critical in online courses due to the greater autonomy that students have in these environments. Through SRL strategies, students can adapt their learning behavior timely.

One of the online tech education companies is Udacity, which offers a variety of Nanodegree programs to teach the latest tech skills. To earn their Nanodegree certificate, students need to complete multiple projects during the term, which is usually a three-to-six-month learning period. Typically, the graduation rate is around 34%, which means that seven out of 10 students fail to complete the coursework (average online course completion rate under 5%). Many factors are involved in the high dropout rates since online learners often have busy lives and the whole term is relatively long. One of the factors is a lack of progress and process awareness, which might lead to students missing the opportunity to self-regulate their learning continuously. In other words, lack of awareness might cause students to fail to plan and allocate the necessary effort and time to reach their goals (Jivet, Scheffel, Drachsler, & Specht, 2017).

Previously built dashboards in Udacity targeted internal teams to keep track of key metrics at cohort-level aggregations (e.g., graduation rate). However, to date, Udacity did not have a process in place to feed back to individual students a summary of their online learning behavior based on trace data as awareness support.

The objective of this work is thus to implement an awareness tool leveraging students' trace data on the Udacity platform to help them continuously monitor their learning processes as a basis for the

SRL needed during their online learning. In particular, Udacity could develop a personal weekly study report including various progress indicators.

PERSONAL WEEKLY STUDY REPORT

The personal study report is a snapshot of an individual learning dashboard, which Udacity sends to students on a weekly basis. The report is divided into three parts, namely, Overall Progress, Study Time, and Project/Concept Completion. The Overall Progress part give students the ratio of the number of days passed since the term start and the length of the terms and the ratio of the number of projects passed versus total. Both ratios are provided in absolute numbers and in percentages. Thus, a quick glance at the two percentages, which are one below the other, gives students the perspective of whether they are on track or behind. This part also presents a chart of cumulative days of study by month. The Study Time part displays the time students devoted to studying during the week, and the cumulative since the term start. The Project/Concept Completion part shows the cumulative projects submitted versus passed, as well as the cumulative number of concepts learned. All the charts on the report present the student's indicators side by side with those average of the cohort and those of the top of the cohort. All the charts involving time enable students to keep track of the effort they have invested in terms of time. The student could then decide whether the effort is enough, or the time allocation should be revised to adapt the effort accordingly. The report is mostly visual and simple, thus allowing students to monitor their progress at a glance. Based on the monitoring results, students could activate SRL strategies to succeed in the course.



Figure 1: Weekly report of one student of the Data Analyst Nanodegree program

The report could be enhanced by adding machine learning algorithms to predict students' likelihood of graduation based on their current situation. The algorithm could be trained using data from past editions of the course. In addition, conducting a study in an experimental setup with a control group without the report and an experimental group with the report, could help to determine the actual impact of the report on students' choice of SRL strategies.

REFERENCES

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice. In L. É., H. Drachsler, V. K. B. J., & P.-S. Mar (Eds.), *Proceedings of the 12th European Conference on Technology Enhanced Learning* (pp. 82–96). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-66610-5_7

The use of an AI algorithm to verify exam questions on prescriptions within e-learning programme P-scribe

Adriaan B. D. van Doorn

Dept. of Clinical Pharmacy and Pharmacology
University Medical Center Groningen
a.b.d.van.doorn@umcg.nl

Floor van Rosse PhD

Erasmus Medical Center Rotterdam, dept. of hospital pharmacy
f.vanrosse@erasmusmc.nl

Students*

Nino Jansen (ninojansen5@gmail.com), **Sofie Lovdal**

(s.s.lovdal@students.rug.nl) **Thomas van Dongen** (thomas123@live.nl)

*Faculty of Sciences and Engineering, Artificial Intelligence and Computer Science, University of Groningen

ABSTRACT: In medical curricula, students learn how to safely and effectively prescribe drugs to (fake) patients. P-scribe (<https://www.pscribe.eu/>) is an e-Learning platform in which students can practice this important skill. P-scribe is used in all medical schools in the Netherlands. In the Erasmus Medical Center the exam module is extensively used to actually examine the prescribing skills of students. Verifying the exam questions is a pretty intensive procedure but as method it can be used to develop an AI algorithm AI/A. “Grading” a prescription consist of several ‘layers’. Obviously a correct drug must be prescribed, in the right admission form, strength, dose, dose frequency, and important dosing information. At this moment a concept version of an AI/A based program has been developed to validate and grade prescriptions. At the LAK conference this tool will be presented. In the coming year we fill P-scribe with various scripts and train the AI/A tool with big data. The grading output of this AI/A tool is compared with teacher-based grades. Our AI/A future is very ambitious.

Keywords: Medication safety, prescribing skills, e-Learning, AI algorithm, automatic grading

1 INTRODUCTION

P-scribe (<https://www.pscribe.eu/>) is an e-Learning platform in which students can practice ‘pharmacotherapeutic reasoning’ using the WHO 6-STEP method; by using the 6-STEP students should come to a safe, effective, and reasonable pharmacotherapeutic treatment. In step 5, an actual prescription is made. Pscribe contains a “case module” and an “exam” module; the first is intended as self-study part and cases in this part are often used in blended education – students can prepare a case before entering the classroom. The exam module is used to examin students (either ‘formative’ or ‘summative’)

2 METHODS

Grading prescribing skills In the Erasmus Medical Center the exam module is extensively used to actually examine the prescribing skills of students. Nowadays, student's prescriptions ($N > 2000$ a year) are verified by teachers using a protocol. Each prescription can get a certain amount of points, and for mistakes/omissions students 'lose' points. Prescription errors with high risk for patient safety (such as a toxic overdose) cost more points; than prescription errors that will for example cost the pharmacy a little bit more time but will not form a risk for patient safety (such as forget to indicate the amount of tablets that should be delivered). Verifying the exam questions is an intensive procedure but as method it can be used to develop an AI algorithm AI/A (2). As already indicated above, "Grading" a prescription consist of several 'layers'. Sometimes, >1 drugs can be prescribed; sometimes one particular drug is the 'best choice' while other drugs are second choice but still deserve some 'points'. At this moment a concept version of an AI/A based program has been developed to validate and grade prescriptions. The teacher can choose the grading scheme he or she wants to use and can weigh the different layers as in some exam questions, choosing the right admission form is important (e.g. a young child which cannot swallow tablets); while in others the right dose deserves a higher weight (e.g. drugs with a small therapeutic window). This means that the tool will not only save time of teachers; but will also allow for a more precise grading than in the old way. Obviously, grading will not become completely automatic, but the obvious pass and obvious fail results will – after extensive validation- be automatically graded while the prescriptions that are borderline pass/fail will still be looked at by a teacher.

3 NEXT STEPS

At the LAK conference this tool will be presented. At this moment, several 'simple' prescriptions can be graded. In the coming year we fill P-scribe with various scripts and train the AI/A tool with big data. The grading output of this AI/A tool is compared with teacher-based grades, and feedback of teachers who typically grade will be taken into account in further development. Our AI/A future is very ambitious. We aim for a broader application of this tool as P-scribe is extensively used in all medical and pharmaceutical schools in the Netherlands, but also outside the Netherlands.

REFERENCES

- Fast, E., Horvitz, E., Long-Term Trends in the Public Perception of Artificial Intelligence. Proceedings of the 31th AAAI Conference on Artificial Intelligence (AAAI-17), 2017, 963-969.
- de Vries TPGM, Henning RH, Hogerzeil HV, Fresle DA. Guide to Good Prescribing. A Practical Manual. Geneva: World Health Organization Action Programme on Essential Drugs; 1994. [30 March, 2006]. WHO/DAP/94.11. http://whqlibdoc.who.int/hq/1994/WHO_DAP_94.11.pdf.
- Doorn, A.B.D. van et al, PScriBE: A Pharmacotherapy e-learning web-application enabling registration and mapping of the rational drug-choice process of students and experts. Basic & Clinical Pharmacology & Toxicology. 2015; Volume 8, Supplement S, E46-E46.

CoTrack: A Tool for Tracking Collaboration Across Physical and Digital Spaces in Collocated Blended Settings

Pankaj Chejara, Luis P. Prieto, María Jesús Rodríguez-Triana, Shashi Kant Shankar

Tallinn University

pankajch@tlu.ee, lprisan@tlu.ee, mjrt@tlu.ee, shashik@tlu.ee

Adolfo Ruiz-Calleja

University of Valladolid

adolfo@gsic.uva.es

ABSTRACT: Collocated collaboration in blended settings involves the usage of technology in addition to face-to-face interactions among participants (Martinez-Maldonado et al., 2017), enabling interactions across physical and digital spaces. LA solutions often rely only on interactions captured in the digital space, offering a partial picture of the learning behavior (Pardo & Delgado Kloos, 2011). Aiming to overcome these limitations, CoTrack- has been developed to offer teachers more holistic information about student's collaboration behavior across physical and digital spaces. CoTrack is based on Raspberry Pi and contains two components for data collection purposes: one for tracking students' digital activity in a real-time collaborative editor (Etherpad), and another for collecting privacy-preserving audio data (it only collects the sounds' direction of arrival, DOA). NTP (Network Time Protocol) is used for handling the time synchronization among the data collected from multiple students' groups, and MQTT (Message Queuing Telemetry Transport) is used for transmitting audio DOA data to a server in real-time. The data collected is then analyzed to build an interaction network that shows "who talks to whom" relationship among students. It was also analyzed to obtain statistics about speaking time and writing activities. Finally, we developed an interactive dashboard for teachers to visualize the data. These features are shown for the entire duration of the collaboration activity and also allow the teacher to change the visualization for different time-frames (e.g. 30 sec, 60 sec, 5 min, etc). CoTrack has been used in an authentic classroom settings involving 1 teacher, 2 researchers and 9 students arranged in groups of 3. We faced the data quality issues because of the classroom noise and students' movement. However, according to the teacher, CoTrack could be helpful for assessing collaborative activities.

Demonstration Movie: <https://www.youtube.com/watch?v=9xmeXmp7Hp8>

Keywords: Collocated collaboration, Multimodal Learning Analytics, Computer-Supported Collaborative Learning

REFERENCES

- Martinez-Maldonado, R., Kay, J., Buckingham Shum, S., & Yacef, K. (2017). Collocated Collaboration Analytics: Principles and Dilemmas for Mining Multimodal Interaction Data. *Human-Computer Interaction*, 34(1), 1-50.
- Pardo, A., & Delgado Kloos, C. (2011). Stepping out of the box: towards analytics outside the learning management system. International Conference on Learning Analytics and Knowledge 2011 (pp. 163-167). ACM, New York.

Creating an Interface Supported by Learning Analytics to Assist Learners with Navigating Individualized Learning Pathways

Matt Crosslin

University of Texas at Arlington

matt@uta.edu

ABSTRACT: What if natural language processing and process modeling could be utilized to help learners create their own learning pathway when given multiple modality options? This **interactive demo** will look at a current research into utilizing chatbots and H5P to create interactive content (Crosslin, 2019). The ultimate goal will be to replace the chatbot interface with a dynamically updated interface that changes content and activities based on user preferences and input (Crosslin, 2018b). Guiding this interface will be learning tactics and strategies determined by a process noted by Matcha, Gašević, Uzir, Jovanović, and Pardo (2019) that utilizes an Expectation Maximization algorithm to cluster sessions generated by a First Order Markov Model. These tactics and strategies will be utilized to guide learners through various pre-determined and open-ended. The theoretical design underlying this design is Self-Mapped Learning Pathways (Crosslin, 2018a), a design methodology in which learners are given one modality that was designed by the instructor and one modality that is self-determined. Learners are allowed to choose either modality or create a custom mix of both at any time. The interface for this methodology is designed to facilitate those choices, and early demonstration versions are being tested in various courses (Crosslin, 2019).

Keywords: Heutagogy, Learning Pathways, Natural Language Processing, Process Mining, Chatbots, H5P, Learning Tactics

1 DEMO VIDEO

The demo video for this session can be found at <https://youtu.be/-fSgnaOlx4I>.

REFERENCES

- Crosslin, M. (2018a). Exploring self-regulated learning choices in a customizable learning pathway course. *Australasian Journal of Educational Technology*, 34(1), 131-144.
- Crosslin, M. (2018b, December 5). What if we could connect interactive content like H5P to artificial intelligence? [Blog post]. Retrieved from <https://www.edugeekjournal.com/2018/12/05/what-if-we-could-connect-interactive-content-like-h5p-to-artificial-intelligence/>
- Crosslin, M. (2019, September 4). Building a self-mapped learning pathways micro-lesson: H5P vs Twine [Blog post]. Retrieved from <https://www.edugeekjournal.com/2019/09/04/building-a-self-mapped-learning-pathways-micro-lesson-h5p-vs-twine/>
- Matcha, W., Gašević, D., Uzir, N. A. A., Jovanović, J., & Pardo, A. (2019, March). Analytics of learning strategies: Associations with academic performance and feedback. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 461-470). ACM.

Student advising learning dashboards: the story of LISSA and LALA

Tinne De Laet, Tom Broos, Martijn Millecamp, Katrien Verbert

KU Leuven, Belgium
first.lastname@kuleuven.be

Julio Guerra

Universidad Austral, Valdivia, Chile
jguerra@inf.uach.cl

Margarita Ortíz

Escuela Superior Politécnica del Litoral, Guayaquil, Ecuador
margarita.ortiz@cti.espol.edu.ec

Miguel Angel Zuniga-Pietro

Universidad de Cuenca, Ecuador
miguel.zunigap@ucuenca.edu.ec

ABSTRACT: Student advising and counseling (Sharkin, 2004) is an essential part of student support during the first years of higher education. The actual advising practices differ a lot between institutions, both regarding the methods used, people involved, embedding in university practices, and maturity. Data-supported advising has the potential of increasing the advising quality and the support provided to individual advisors, and therefore has gained interest (Stoneham, 2015).

Within the LALA project, the LISSA dashboard (Charleer et al., 2018) that supports the interaction between students and student advisors, successfully developed by and deployed at KU Leuven in Belgium, was adapted and adopted in three Latin American universities (Cuenca and ESPOL in Ecuador and Austral in Chile). In the interactive demo the following aspects will be covered;

- the adaption and adoption process of the LISSA dashboard, and how this was influenced by the local context,
- specifics of the institutional dashboards with focus on modules related to drop-out prediction, peer-comparison, course registration, ...
- evaluation of the dashboards regarding use, perceived usefulness, and impact.

The demo will provide ample opportunities for discussion, involving different kinds of stakeholders ranging from policy makers, practitioners to researchers.

Keywords: learning dashboards, student advising, case study, at scale, case study

Video: https://kuleuven.mediaspace.kaltura.com/media/The+story+of+LISSA+and+LALA/1_289q33ci

REFERENCES

- Sharkin, B. S. (2004) College counseling and student retention: Research findings and implications for counseling centers, *J. College Counseling*, 7(2), 99-108
- Stoneham, R. (2015), Failing students need big data and learning analytics: Hype or reality?", *Compass: J. Learn. Teach.*, 7 (11)
- Charleer, S., Moere, A. V., Klerkx, J., Verbert, K., and De Laet, T. (2018). Learning Analytics Dashboards to Support Adviser-Student Dialogue, *IEEE Transactions on Learning Technologies*, 11(3), 389-399, <http://dx.doi.org/10.1109/TLT.2017.2720670>

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

PerformanceVis: Homework & Exam Analytics Dashboard for Inclusive Student Success

Xiaojing Duan, Alex Ambrose, Chaoli Wang, Kevin Abbott, Victoria Woodard, Catlin Schalk

University of Notre Dame

xduan@nd.edu, gambrose@nd.edu, chaoli.wang@nd.edu, kabbott@nd.edu,
vweber1@nd.edu, cschalk@nd.edu

ABSTRACT: PerformanceVis is a visual analytics tool developed for analyzing and visualizing students' chemistry course performance through the lens of time, homework and exams, and demographic and academic background. The Introduction to Chemical Principles course is a required course for all college of science and college of engineering programs at the university and is the second largest course on campus with approximately 1,000 freshmen taking the course. This engaging tool includes four main views (overall exam grade pathway, detailed exam grade pathway, detailed exam item analysis, and overall exam & homework analysis) that are dynamically linked together for user interaction and exploration. PerformanceVis enables instructors to improve their course and assessment design by visualizing students' perceived difficulty level and topic correlation between assignments and exams. It assists instructors and administrators in evaluating the impact of a special treatment program (cohort) by reviewing the performance of regular, control, and cohort students overall and by exam. The right side of Figure 1 shows a view of the gender performance gap for those students who scored a C or below course grade. The left side of Figure 1 shows Exam 1 item analysis for each test question.

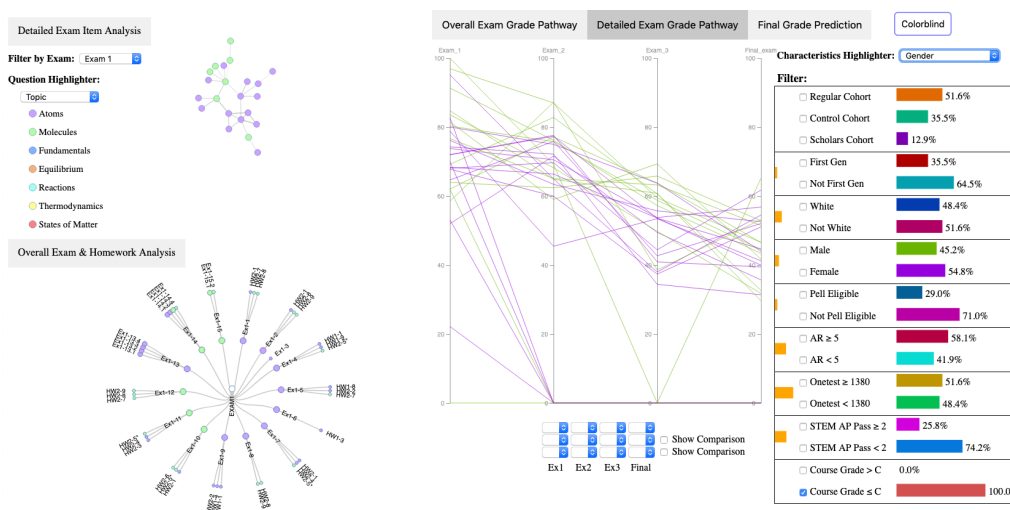


Figure 1: PerformanceVis: Homework & Exam Analytics Dashboard Screenshot

Link to the interactive demo video: <https://youtu.be/5ub7BxYbb5k>

MINDSTEPS: An Adaptive Computer-Based Tool for Formative Student Assessment

Nina König

Institute for Educational Evaluation, Associated Institute at the University of Zurich
nina.koenig@ibe.uzh.ch

Martin J. Tomasik, Stéphanie Berger, Lukas Giesinger, Laura A. Helbling, and Urs Moser

Institute for Educational Evaluation, Associated Institute at the University of Zurich

ABSTRACT: We demonstrate the application of a computer-based formative assessment tool that is currently used in several cantons of Switzerland. MINDSTEPS can be accessed anytime via computer or tablet. A distinctive feature of it is that it covers topics and competencies from third grade in primary school through third grade in secondary school, spanning seven years of compulsory schooling. It is a tool to systematically assess competencies defined in the official competency-oriented curriculum of the German speaking cantons in Switzerland. The underlying item bank currently contains between 4,000 and 12,000 items per subject. There are two thematically identical types of item banks: a practice and a testing item bank. The former is openly available to students for training. Students can autonomously create assessments from a topic domain and difficulty level they choose. The latter is for teachers to evaluate students' learning progress and to identify their strengths and weaknesses. Teachers can select items within desired competencies or curricular topics. In this demonstration, we will show how MINDSTEPS is used by students and teachers.

Keywords: assessment for learning; formative student assessment; feedback; online tool; computer-based learning

Link to video: <https://vimeo.com/362756278> (click on CC for English subtitles)

FLOWer: Feedback Loop for Group Work Supporter

Yoon Lee^{1#}, Haoyu Chen³, Esther Tan¹, Sambit Praharaj², Marcus Specht¹

¹Delft University of Technology, Netherlands, Open University of Netherlands, Netherlands,

³University of Oulu, Finland. #Email: y.lee@tudelft.nl

ABSTRACT: This interactive demo presents FLOWer, a recently developed automated prototype to facilitate and support collaborative learning in group-work settings. The prototype enables the assessment, analytics and support of Computer Supported Collaborative Learning (CSCL) in the real-world scenario with sensor-data from multiple modalities. It analyzes the pre-designed indicators with multimodal learning analytics (MMLA) for providing real-time and metaphorical feedback on individual/group's contribution, to inform teaching and learning practices towards a better interaction loop in educational settings. The overall architecture of the FLOWer consists of three components: 1) multimodal data stream collection, 2) multimodal learning analytic and 3) metaphorical visualization. In component 1), the relevant features are extracted from the raw data of the cameras, audio recorders, and the text. Next, neural-network-based learning methods are utilized to assess the CC with different indicators. At last, the real-time metaphorical feedback (a landscape with components like the plants, wind, rain) from different indicators (i.e., speaking time, specific gestures) will be visualized via a dashboard. The FLOWer serves as an initial implementation of MMLA for CSCL, to facilitate feedback provision and to explore actionable feedback at the individual and group level.

Keywords: computer supported collaborative learning, multimodal learning analysis, neural networks, metaphorical design, real-time interaction

Demo video link: <https://youtube.com/watch?v=pfbISWSznB0>

REFERENCES

- Tausch, S., Hausen, D., Kosan, I., Raltchev, A., & Hussmann, H. (2014). Groupgarden: Supporting Brainstorming through a Metaphorical Group Mirror on Table or Wall. *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*.
- Streng, S., Stegmann, K., Hußmann, H., & Fischer, F. (2009). Metaphor or Diagram? Comparing Different Representations for Group Mirrors. *In Proc. OZCHI, ACM*, (pp. 249–256).

ACKNOWLEDGEMENT

This research is part of a larger project on “Uncovering the Process and Outcome of Computer-Supported-Collaborative-Learning (CSCL) using Multimodal Learning Analytics: New Environments and Assessment for CSCL in the learning continuum” with National Institute of Education (NIE), Nanyang Technological University (NTU), Singapore. We thank the NIE team for their support.

The Fellowship of the Learning Activity: playing, cooperating, creating awareness and designing learning activities

Marcel Schmitz

Zuyd University of Applied Sciences
marcel.schmitz@zuyd.nl

Maren Scheffel

Open University The Netherlands
maren.scheffel@ou.nl

ABSTRACT: In the last years, research to connect learning analytics to learning design has been on the rise, but there still are a number of steps that need to be taken in order to make a workable connection between learning analytics and learning design. To improve the quality it is important to capture and structure the design choices and to retrieve data on the behavior, the effects and the opinions about the designed learning activities. In an effort to get (1) input on the learning design choices of learning activities and (2) bridge the gap between learning analytics and learning design, a board game has been developed. The Fellowship of the Learning Activity is a serious game that captures and systematizes the learning design of learning activities. Additionally, the game brings awareness for the players of the multidisciplinary approach of learning design and the connection of learning analytics to learning design. The demo shows a play session of The Fellowship of the Ring. It shows the game elements that are used to connect learning analytics and learning design and it shows how learning design choices can be captured and systematized.

Keywords: Learning Analytics, Learning Design, Learning Activities, Boardgame

A video of the demo is available at: <http://somup.com/cqXVYAfiUD>

Berkeley Online Advising (BOA): Learner Analytics as an Advising Resource

Author(s): Steven Williams, MLIS

University of California, Berkeley

stevenwilliams@berkeley.edu

ABSTRACT: Berkeley Online Advising, known as BOA, is a dashboard for undergraduate advisors that unifies student data from a variety of campus sources, including the LMS (Canvas), student information system, and departmental databases, in a highly performant cloud-based data lake architecture. BOA was launched as a campus IT service in July 2019 and is currently available to over 900 advising staff. BOA uses campus and LMS data to help advisors organize their students into data-driven cohorts, and generates real-time alerts when students are not making regular academic progress. By highlighting details about students at greatest academic risk, BOA supports advisors by helping them identify the individual students among their hundreds of advisees who would most benefit from advising interventions and additional support.

BOA also supports advisors in their day-to-day workflow, by providing new tools for advisors to capture notes about their interactions with students. Over 10,000 advising notes were authored using BOA in the first three months of the tool being available campuswide. Pooling these notes into BOA's shared data lake helps provide more evidence and context for future opportunities to explore how student behaviors and advisor support may correlate with student outcomes.

For more information, visit: <https://youtu.be/hNMylKg-1xE>

Keywords: advising, advisors, canvas, LMS, data lake, student information system, SIS, visualization, AWS, student success, student outcomes

A novel feedback system for refinement and improvement of lecture-based pedagogy

Pankaj Chavan

IDP in Educational Technology, IIT Bombay

pankajchavan@iitb.ac.in

ABSTRACT: Despite the best efforts of educators in implementing active learning strategies, most classrooms remain largely lecture driven. In such classrooms, students often cannot voice their opinion on the lecture content and teaching that are likely to affect their learning. At the other side of the lectern, the teacher is also left unaware of the efficacy of her teaching methods and has to rely on summative assessment for this purpose, which provides little insight on the learning process of the students. Student feedback is one of the approaches to provide formative feedback about instruction to teachers, which can help teachers in improving instruction. Hence, this research focuses on designing and developing a unique feedback system which enables students to reflect on their cognitive-affective states in real-time which can then be used by teachers to improve pedagogy in such lectures. The findings from this research will demonstrate how such feedback can identify optimal or sub-optimal lecture content, track cognitive-affective dynamics in a classroom, assist instructors in retrospective self-evaluation and improve instruction. We will also investigate how giving such feedback impact student learning.

Keywords: Student feedback, lecture-based classrooms, cognitive-affective states, live feedback

1 BACKGROUND AND MOTIVATION

Consider this scenario: *Dr. Taylor received a PhD degree from an Ivy League institution. She was quickly inducted to teach a large introductory classroom at a land-grant university. Having a background of high-quality research and teaching, she assumed the responsibility of inculcating in the students a sense of wonder, logic and understanding that are critical for an introductory course. In one of the class, she had to introduce the concept of oxidation and reduction. Dr. Taylor thought she had described the concept very well. However, after the exam, she figured out that most of the students did not understand the concept. In addition to that, she recalled retrospectively, of the few students (<1% of the class size) who had come to see her before the exam, most had trouble with this particular concept. Despite her best intentions of identifying and analyzing the source of the problem, she was first, not able to identify the section(s) of the lecture that was commonly difficult for the class to understand and second, even after getting feedback in the form of students' exam performance, she was uncertain of the changes that needed to be made in the instruction.*

From the teacher's perspective, there is a 'pedagogical blind spot', which if identified, can help her in addressing existing problems in quasi-real-time to resolve student understanding issues and improve the instruction in successive iterations of the same course. One of the solutions for such problematic situation is teacher reflection (Ashwin, Boud, Coate, Hallett, & Keane, 2015). Reflection initiates teacher questioning and thinking about their teaching practices, thus resulting in more informed

teacher actions which benefit both the teacher as well as students. Evidence from the classroom (e.g. student performance, student feedback of teaching etc.) is considered as crucial for promoting teacher reflection (Ashwin et al., 2015; Pollard, Black-Hawkins, Cliff-Hodges, Dudley, & James, 2014) because it enhances the quality of teacher's judgement and decisions about assessing and improving the instruction. However, in the above-mentioned situation, the evidence, i.e. poor exam performance, is not sufficient to identify the 'pedagogical blind spot' and critically reflect upon the aspects of the instruction to evaluate and improve it further. There are two critical issues here: first, challenges associated with generating evidence at a more granular level from the classroom for evaluating and improving instruction, and second, lack of formal training to teachers in higher education on practicing reflective approach of teaching (Ashwin et al., 2015).

Now, let us consider the same problem from a coexisting perspective; that of the student. From the student's perspective, they faced several learning obstacles during that lecture due to which the instruction failed to address pre-existing knowledge gaps or created new ones. If such gaps can be identified or corrected early on in a career, then it can make future learning easier and meaningful, because it is well recognized that new knowledge is constructed based upon existing knowledge and beliefs (Bransford, Brown, & Cocking, 1999). In-class participation of students, which consists of questioning, raising hands and making comments (Rocca, 2010), can help students to circumvent the learning obstacles. However, students in college classrooms often fail to participate in class (Rocca, 2010) due to several individual and classroom-related factors such as student and instructor gender, class size etc., which can critically influence student learning and potentially hinder their understanding of key concepts. Additionally, due to difficult learning materials, the students got disengaged and they did not find any value and relevance in learning the material resulting in reduced attention to the lecture materials. Hence, in such situations, there is a need for a mechanism to scaffold the students to sustain their attention on classroom learning in order to positively affect academic performance (Wei, Wang, & Klausner, 2012). Thus, in such situations, there is a need for a mechanism to address above mentioned issues and support student learning.

Therefore, we are faced with a scenario where despite the ubiquity of lecture-based classrooms (Stains et al., 2018) sub-optimality of lectures are a cause of concern from both the teaching and learning perspective and one which still lacks a viable solution. This is the motivation of our research: *To understand and establish the value of using real-time student feedback in lecture-based classrooms to assess and improve instruction. Furthermore, use this understanding to design a technology solution to assist teachers in improvement of instruction and scaffold student learning.*

2 CONCEPTUAL FRAMEWORK

Broadly, our research is based on the three key research areas in teaching and learning: role of emotions in learning, student feedback for assessing and improving instruction, and reflective practice of teaching. In the last two decades, understanding the role of emotions in learning has continuously been the focus of educational research (Graesser & D'Mello, 2012; Pekrun, 2006) and the importance of student emotions in learning has been demonstrated by the researchers (Pekrun, 2006). Learning is a multidimensional process where cognition, emotion, and motivation come together to produce the end result. There are several theories/frameworks that discuss how these factors combine to produce learning. According to Graesser & D'Mello (2012), students experience several emotions when they are assigned a difficult material to learn or difficult problems to solve,

such as confusion, frustration, boredom, and engagement/flow, which play a crucial role in student learning. According to Pekrun's (2006) control-value theory, the student emotions (e.g., boredom) influence available cognitive resources during learning (e.g. attention in a given academic activity). Csikszentmihalyi's (1990) flow theory proposes that if the difficulty of the task is too low as compared to any individual's skills then the individual experiences boredom for a task, whereas anxiety is experienced when the task difficulty is too high as compared to an individual's skills. The theories/frameworks mentioned above demonstrate that cognition and emotion are tightly coupled. Hence, we see value in harvesting the real-time data about student's cognitive-affective states.

Student feedback is appreciated as a key source of evidence for assessing and improving teaching quality (Richardson, 2005). However, the literature on student feedback highlights that most of the research followed a top-down approach for collecting the student feedback to assess and improve instruction. For example, student evaluations of teaching (SETs) (Richardson, 2005) ask students to provide ratings on predefined dimensions of teaching. However, in case of clickers, though the student feedback is captured in real-time, it is instructor initiated and discontinuous (i.e. collected at discrete intervals of time). Hence, it may miss out some important information. There are some exceptions where the bottom-up approach has been used to gather feedback, such as Google-glass based classroom feedback system (Zarraonandia, Díaz, Montero, Aedo, & Onorati, 2019) and Live Interest Meter (LIM) (Rivera-Pelayo, Munk, Zacharias, & Braun, 2013). But, the data these systems collect is unidimensional (e.g. confusion level or comprehension) and is incapable of capturing the multidimensionality of the learning process. Additionally, the student's perceptions of their learning environment have been identified to influence their learning, both positively or negatively (Fraser, 2012). Hence, we propose a bottom-up approach where student-initiated, real-time feedback on their cognitive-affective states is collected to understand the exact nature of problems students face in a classroom at a higher granularity. The literature on reflective practice of teaching suggests that through the process of reflection, teachers become self-aware that problem exists in their teaching practices. Teachers view the problematic situations from different perspectives, question their own actions and teaching practices and take decisions related to the future plan of actions (Zeichner & Liston, 1987). It is considered to help students learn in meaningful ways. Hence, we also focus on providing the feedback data and associated resources to teachers to facilitate reflection, and hence improve lecture-based classroom instruction.

3 RESEARCH GOAL AND QUESTIONS

Our broad research goal is to understand and establish the potential of real-time student feedback about their cognitive-affective states for assessing and improving instruction and designing a technology-supported feedback system for improving lecture-based classroom instruction through teacher reflection. I argue improvement of student learning will be an important by-product of the proposed technology. Following are the specific research questions:

- RQ1: How is the feedback data useful in fine-tuning the lecture-based classroom instruction?
- RQ2: What are the student's and teacher's perceptions of the usability and usefulness of the feedback system?
- RQ3: What cognitive-affective dynamics exists in lecture-based classrooms?
- RQ4: How is the feedback data useful as evidence to promote teacher reflection?
- RQ5: What is the impact of giving such feedback on student learning?

4 METHODOLOGY, CURRENT STATUS AND RESULTS

Our research problem has the following requirements: the necessity of testing and refining the design in a real-world setting to develop plausible solutions, the need of different research methods that can cater to different phases of the research, and the requirement of evaluating effects of features of the proposed solution on teaching practices and student learning. The requirements of our research problem align well with the characteristics of design-based research (DBR) (Wang & Hannafin, 2005). Hence, DBR was chosen as the methodology for our research. We implemented the initial prototype of the feedback system in the lab as well as a classroom setting, and we consider this as the first cycle of DBR.

- *Phase one: Problem analysis based on the synthesis of literature and existing solutions*

Initially, the research problem was determined and refined further through the synthesis of literature on three key research areas: role of emotions in learning, student feedback for assessing and improving instruction, and reflective practice of teaching (See Section 2). We also identified problems associated with the potential existing solutions which can address the research problem we are looking at, such as teacher-initiated and unidimensional nature of feedback from students (Chavan & Mitra, 2019a). Then, the results of the synthesis of literature were combined with the problems identified with the existing solutions to propose the initial solution.

- *Phase two: Initial solution design informed by existing design principles*

We developed the initial prototype of the feedback system interface (See Figure 1 (Left) in Appendix) based on the following two design considerations: 1. Learning is a multidimensional process - Keep complementary cognitive-affective states, i.e. both positive and negative, and 2. The feedback system should be as unobtrusive as possible for the students (details in Chavan & Mitra, 2019a). The simple web-based application (Figure 1 (Left) in Appendix) collects **anonymous** and **continuous** feedback from students on four variables, namely, easy, difficult, engaging and boring, to capture their cognitive-affective states in the classroom. The other decision made was about videotaping the lecture. We decided to videotape the lecture and sync it with the student feedback for retrospective analysis of the instruction.

- *Phase three: Evaluation and reflection*

The first study with the intervention was conducted in a lab-based setting with 10 students. The results revealed that the feedback system was able to capture certain perceptions of the lecture at high granularity (peaks of difficulty and engagement in certain sections) which would otherwise have remained obscure. A usability survey indicated a positive impression of the feedback system in general. These initial results of the study were presented at the ICCE 2018 conference as a short paper (Chavan, Gupta, & Mitra, 2018). Interviews with students and instructor revealed a difference between the students' and the instructor's perception of the probable causes of a dominant cognitive/affective state (i.e. peaks of difficulty and engagement) (See Appendix Figure 2). Understanding such differences in perceptions can act as evidence for the instructors to reflect on the instruction and gain deeper insights about the potential source of the problem. These results have been reported in a manuscript in preparation (Chavan & Mitra, 2019b).

The second study was conducted in a lecture-based classroom with 30 participants, in which feedback data was collected in 18 lectures. The data collected during first 5 lectures of the second

study was reported in a short paper presented at LAK 2019 (Mitra & Chavan, 2019), which aimed at exploring the pedagogical affordances of the proposed feedback data. The detailed results of the second study (based on data collected in 13 lectures) will be presented at the T4E 2019 conference, which is accepted as a full paper (Chavan & Mitra, 2019a). The feedback data demonstrated how such data could be useful for instructors to self-reflect on their instruction. The focus group interviews demonstrated that students perceived the feedback system useful for teachers to assess and improve instruction. Students perceived value in notifying student feedback to instructors for the appropriate action to be taken in the next class to resolve the problems.

Based on the results of the two studies (Chavan & Mitra, 2019a, 2019b), we proposed some modifications in the design and implementation of the feedback system. The major decision we have taken is closing the loop, i.e. channelling the feedback along with student reasoning to the teacher to address the problems in quasi-real-time. The system which is redesigned based on this decision will have two components: one, the student feedback interface (Figure 1 (right)) and, second a simple web-based teacher dashboard with different visualizations (Figure 3), which acts as an evidence for assessing and refining instruction through reflection. The modified intervention will be tested again with the same steps of DBR, which will form the second cycle of DBR.

5 NEXT STEPS

The results of the first iteration of feedback system evaluation demonstrated the potential of the feedback data to be used as evidence to assess and improve classroom instruction through reflection. It has also demonstrated the ability of feedback data to explore classroom cognitive-affective dynamics (Mitra & Chavan, 2019). Such analysis has two-fold potential: first, it can inform teaching, and second, such subtle dynamics if explicated can inform new learning theories and advance old ones. For example, how the flow theory which is applicable to individuals can apply equally well to classrooms as a whole. The focus group interview results reported in (Chavan & Mitra, 2019a) also demonstrated how such feedback is capable of capturing and demonstrating the link between cognition and affect in a classroom setting. Example excerpt: *".....We click on difficult if we are not at all able to understand the [content] taught. And if the lecture continues like that [difficult], then we click on boring"*. In the next iteration of the evaluation studies, we will be exploring **RQ3**. We also want to explore the potential of such feedback data to act as a source of evidence for classroom enquiry, thus informing the teacher's judgments and decisions about revising the teaching practices. We will be looking at this aspect (**RQ4**) through the theoretical lens of the Pollard's model of reflective teaching because this model views evidence from the classroom as crucial in the reflective practice of teaching.

The focus group interview results reported in (Chavan & Mitra, 2019a), also demonstrated how such feedback is capable of scaffolding metacognition. Example excerpt: *"...in the starting 5 minutes I just take time to understand...I take time to settle myself. Then, I like...umm...try to analyze whether I am able to understand or not. If I am not able to understand, then I click."* Here, the student is reflecting on his experience during learning in order to provide feedback about his cognitive state. Such initial results strengthen our conjecture that providing such feedback will improve the metacognitive experiences of students. Additionally, when a student provides feedback, s/he will cogitate on the aspects of his own learning and retain sustained attention to lecture materials, which could lead to better learning from lectures. We will be exploring both the aspects, i.e. learning performance and

metacognition (**RQ5**), in the next iteration of evaluation. The student's perception of usability and usefulness of the initial prototype has been positive. However, closer collaboration with the teacher is needed to understand their perceptions of usability and usefulness of the student feedback and the proposed teacher dashboard (**RQ2**).

REFERENCES

- Ashwin, P., Boud, D., Coate, K., Hallett, F., & Keane, E. (2015). *Reflective teaching in higher education*. Bloomsbury Publishing.
- Bransford, J. D., Brown, A., & Cocking, R. (1999). *How people learn: Mind, brain, experience, and school*. Washington, DC: National Research Council.
- Chavan, P., Gupta, S., & Mitra, R. (2018). A novel feedback system for pedagogy refinement in large lecture classrooms. *International Conference on Computers in Education*, 464–469. Philippines.
- Chavan, P., & Mitra, R. (2019a). Developing a student feedback system using a design-based research approach. (In Press) *Proceedings of the 10th International Conference on Technology for Education (T4E)*.
- Chavan, P., & Mitra, R. (2019b). Multiscalar nature of student-teacher perceptions of lecture content. *Manuscript in Prep*.
- Csikszentmihalyi, M. (1990). *Flow. The Psychology of Optimal Experience*. New York (HarperPerennial) 1990.
- Fraser, B. J. (2012). Classroom learning environments: Retrospect, context and prospect. In *Second international handbook of science education* (pp. 1191–1239). Springer.
- Graesser, A. C., & D'Mello, S. (2012). Emotions during the learning of difficult material. In *Psychology of learning and motivation* (Vol. 57, pp. 183–225). Elsevier.
- Mitra, R., & Chavan, P. (2019). DEBE feedback for large lecture classroom analytics. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 426–430. ACM.
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18(4), 315–341.
- Pollard, A., Black-Hawkins, K., Cliff-Hodges, G., Dudley, P., & James, M. (2014). *Reflective Teaching in Schools: Evidence-Informed Professional Practice*. Bloomsbury Publishing.
- Richardson, J. T. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment & Evaluation in Higher Education*, 30(4), 387–415.
- Rivera-Pelayo, V., Munk, J., Zacharias, V., & Braun, S. (2013). Live interest meter: Learning from quantified feedback in mass lectures. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 23–27. ACM.
- Rocca, K. A. (2010). Student participation in the college classroom: An extended multidisciplinary literature review. *Communication Education*, 59(2), 185–213.
- Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., Laski, F. A. (2018). Anatomy of STEM teaching in North American universities. *Science*, 359(6383), 1468–1470.
- Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, 53(4), 5–23. <https://doi.org/10.1007/BF02504682>
- Wei, F.-Y. F., Wang, Y. K., & Klausner, M. (2012). Rethinking college students' self-regulation and sustained attention: Does text messaging during class influence cognitive learning? *Communication Education*, 61(3), 185–204.
- Zarraonandia, T., Díaz, P., Montero, Á., Aedo, I., & Onorati, T. (2019). Using a Google Glass-based Classroom Feedback System to improve students to teacher communication. *IEEE Access*, 7, 16837–16846.
- Zeichner, K., & Liston, D. (1987). Teaching student teachers to reflect. *Harvard Educational Review*, 57(1), 23–49.

APPENDIX

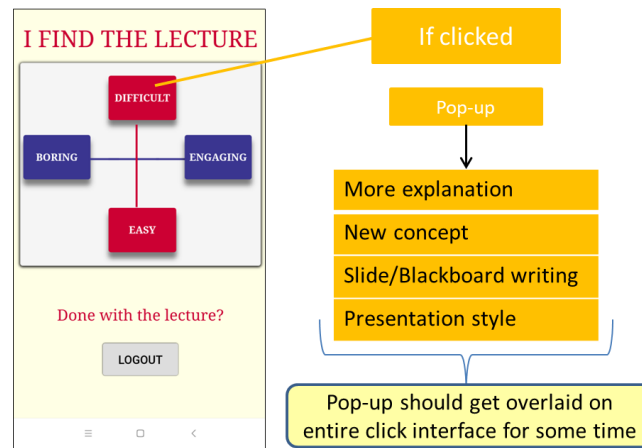


Figure 1: (Left) An android phone view of the web application used in the study. Students can give feedback about their cognitive-affective states by clicking on the four buttons. (Right) A proposed addition to the feedback system interface

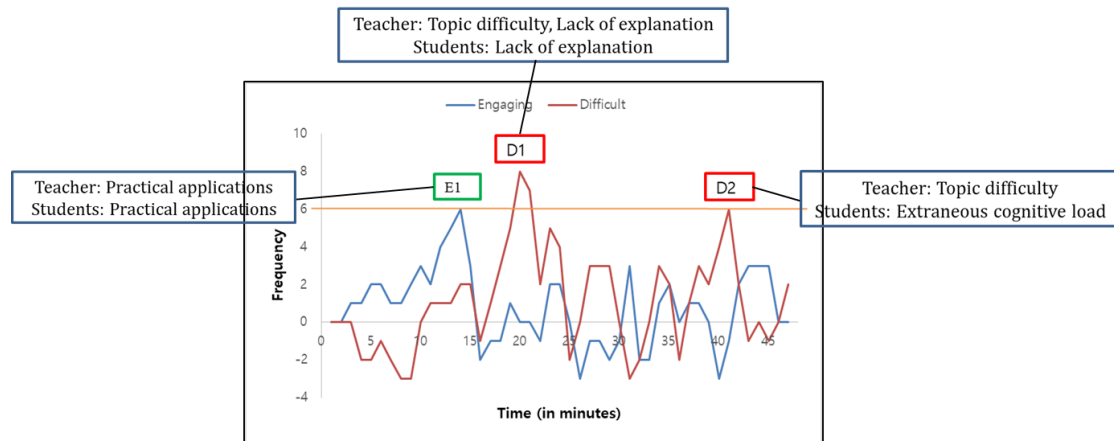


Figure 2: Students and instructor differed on their perceptions of the causes of peaks observed in the data. Boxes contain the primary reasons put forward by the two groups when retrospecting

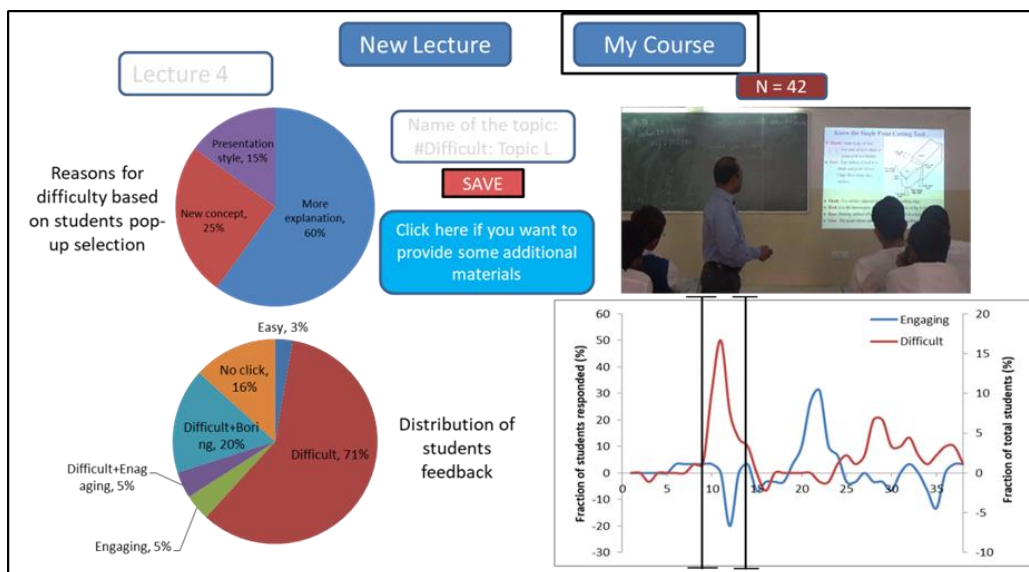


Figure 3: Proposed design of a teacher dashboard with student feedback visualizations

Modelling Student Participation Using Discussion Forum Data

Elaine Farrow

School of Informatics, University of Edinburgh, Edinburgh, UK
Elaine.Farrow@ed.ac.uk

ABSTRACT: Across many different educational settings, course discussion forums allow students to learn from one another and connect socially with their peers and instructors. Content analysis of the messages that are exchanged has been used to model engagement using two well-established theoretical frameworks, Community of Inquiry and ICAP. However, manual content analysis is slow and expensive, and prior work on automation is limited. In addition, these two theoretical frameworks developed out of different disciplines, and little work has been done to bring them together. To address these issues, I will evaluate the use of advanced methods from natural language processing to automate the content analysis, considering both frameworks individually and together, and comparing the results with prior work in terms of accuracy and explanatory power. I will also contribute to the conceptual understanding of what characterises a high quality discussion forum contribution by identifying connections between the frameworks themselves and places where they offer complementary perspectives.

Keywords: learning analytics, student engagement, Community of Inquiry, ICAP, natural language processing

1 INTRODUCTION

Course discussion forums are increasingly used to support large face-to-face classes, in addition to their on-going key role in online and distance learning courses. However, the volume of messages exchanged is often so great that instructors can struggle to read them all in a timely manner, or to identify common themes and threads of argument between them. These messages provide a rich source of material for researchers interested in studying how effective learning takes place through discussion (Garrison, Anderson, & Archer, 1999), and there is growing interest in using this data to create models of student engagement. Content analysis techniques can be used to classify the depth and quality of messages using labels from an educational framework, in order to identify conversation threads that are developing appropriately and those that have stalled or are off-task. Two popular frameworks for modelling student engagement are the Community of Inquiry (Col) framework (Garrison, Anderson, & Archer, 1999), and the ICAP framework (Chi & Wylie, 2014). Col is one of the best-studied theoretical frameworks in online education (Gašević, Adesope, Joksimović, & Kovanović, 2015), and ICAP has been used as a foundation for many studies on computer-supported collaborative learning (Wang, Yang, Wen, Koedinger, & Rosé, 2015). By automating the content analysis, the results can be used while a course is still running. For example, instructors could be notified about conversation threads where they might want to intervene (although the specifics of that intervention are out of scope for this research project). Automation also allows research to be done on large data sets where manual annotation is impractical. Computational models that can assign labels to new data can also provide further insights by revealing patterns within the data. For example, a random forest model can report which of the model features was most informative.

In my doctoral work, I will make use of both the Col and ICAP frameworks and evaluate the ability of advanced methods from Natural Language Processing (NLP) to automate the content analysis based on the labelling schemes provided by the two frameworks. My goal is to improve the way we identify and model the depth and quality of student participation using discussion forum data. I aim to develop methods that handle input text more flexibly, while producing outputs that are at least as accurate and informative as previous work. My work will also contribute to a better conceptual understanding of engagement through analysis of the relationship between the frameworks.

2 BACKGROUND

2.1 Theoretical frameworks for modelling student engagement

2.1.1 *Community of Inquiry (Col)*

The Community of Inquiry (Col) framework for online education is a powerful tool for analysing and developing effective learning experiences (Garrison, Anderson, & Archer, 1999). The framework identifies three main elements ('presences') that are important for a successful educational experience: i) a social environment conducive to learning (*social presence*); ii) a well-designed course with on-going facilitation (*teaching presence*); and iii) the student's own cognitive engagement with the subject matter (*cognitive presence*). Col has been widely used to analyse student learning in online courses (Gašević, Adesope, Joksimović, & Kovanović, 2015), and predictive models have been developed for identifying its elements automatically using the text of discussion forum messages (e.g., Waters, Kovanović, Kitto, & Gašević, 2015).

Two recent studies (Kovanović, et al., 2016; Neto, et al., 2018) that developed models for predicting the phases of cognitive presence both reported high accuracy scores for the prediction task, using linguistically-motivated features as input to the model – things like text coherence, complexity, and readability scores. These were derived from the messages using the text analysis tools LIWC (Linguistic Inquiry and Word Count) (Tausczik & Pennebaker, 2010) and Coh-Metrix (McNamara, Graesser, McCarthy, & Cai, 2014). The features were chosen because they have potential explanatory power, and the studies explored which of them were most predictive. However, the value of the feature analysis is called into question by doubts surrounding the validity of the models themselves. A replication study (Farrow, Moore, & Gašević, 2019) showed that data contamination between the training and testing data in these studies could have led to over-optimistic accuracy scores. Furthermore, only 9 of the top 20 most predictive features from one study (Kovanović, et al., 2016) were still in the top 20 after avoiding the potential contamination, suggesting that over-fitting may have led the prior model to see some features as more predictive than was really the case, and to disregard others that actually have more discriminative power (Farrow, 2018). Therefore, further investigation is needed into the features that characterise high quality discussion contributions.

2.1.2 *ICAP*

The ICAP framework (Chi & Wylie, 2014) defines cognitive engagement based on overt behaviours alone. The framework looks at individual learning activities and how they relate to students' cognitive engagement with the learning materials. Four 'modes' of engagement are identified, and the framework predicts that higher modes will be correlated with greater learning gains. The four modes, in descending order, are **I**nteractive, **C**onstructive, **A**ctive, and **P**assive. Each of these modes subsumes the modes below it and represents a qualitatively different *kind* of growth in knowledge,

not simply a bigger change. Passive engagement corresponds to the least taxing on-task activities; for example, listening to a lecture. Active engagement covers activities that demand the student's attention, such as highlighting lecture notes. To qualify as constructive engagement, novel output must be generated – for example, summary notes. Interactive engagement requires interaction with a partner, and normally both partners must be engaged constructively. However, this requirement is relaxed in the case of activities involving larger groups, since subsets of participants can engage with the same task in different ways. Off-task behaviours do not constitute any type of engagement. ICAP has recently been used to classify student comments on MOOC videos (Taskin, Hecking, Hoppe, Dimitrova, & Mitrovic, 2019). Modified versions of ICAP have been used to analyse discussion forum messages in MOOCs (Wang, Wen, & Rosé, 2016) and student comments on an annotated electronic course text (Yogev, Gal, Karger, Facciotti, & Igo, 2018). Future work can build on this foundation.

2.1.3 Comparing the frameworks

While both frameworks address engagement, they do so from different perspectives. They were developed independently and with different goals in mind. Col was developed specifically in order to understand the benefit of online education and to explain how students are able to learn and develop ideas through discussion. ICAP has a broader scope and has been demonstrated to be effective in predicting the educational value of several different interventions, in a classroom setting as well as online. Little prior work has been done to compare the frameworks, either conceptually or through experimentation. If the labels they assign to messages are found to be closely correlated, then results derived using each of them in previous studies can be expected to be applicable to work using the other. If, instead, they are completely distinct, then using them together in future studies will give a richer picture of engagement. A triangulation study involving both conceptual and empirical comparisons of the frameworks would thus offer a useful contribution to the theoretical understanding of online learning, critical discourse, and learning through discussion.

2.2 Neural network models and advanced NLP methods

In recent years, the field of natural language processing has increasingly embraced the use of neural network methods to classify text automatically. State-of-the-art neural networks can be used to produce accurate outputs for many application domains using only text as input, without the need for extensive feature engineering (Goodfellow, Bengio, & Courville, 2016). Many such applications make use of pre-trained language models known as **word embeddings** (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013; Pennington, Socher, & Manning, 2014), which transform words into points in a high-dimensional vector space. In the high-dimensional space, words with similar meanings are found near one another, while dissimilar words are far apart. This means synonyms are treated in similar ways and common spelling errors can be handled automatically. This approach could be very beneficial for automated content analysis of forum messages, which often contain misspellings.

Neural network models can be hard to interpret, but the use of an **attention layer** (Wu, et al., 2016) is often described as allowing researchers to 'see inside' what is otherwise a black-box technique. After transforming each of the input words into a vector and processing those vectors through the early layers of the network, the attention layer combines the results in a weighted sum before passing them on to later layers. The learned weights in the attention layer can thus be seen as indicating the strength of influence of each of the input words on the final output classification; that is, the extent to which each of the words in a message determines the quality of the contribution.

Work on **multi-task and transfer learning** (Collobert & Weston, 2008) has shown that training a single neural network to learn to generate multiple target outputs at the same time can help to avoid over-fitting to the training data and produce better models overall. This suggests that learning the labels for both Col and ICAP together could work better than using either framework alone.

3 AIMS OF THE RESEARCH

The overall problem that my research will address is how we can identify and model the depth and quality of student participation using the messages that students post to course discussion forums. My research has two main goals: 1) to discover where the Col and ICAP frameworks take a similar approach and where they provide complementary insights; and 2) to evaluate the use of advanced NLP methods to automate the labelling process on new data. Specifically, I will investigate the performance of models that use techniques including word embeddings, attention layers, and multi-task and transfer learning. This work aims to answer four specific research questions.

RQ1: What is the association between the phases of cognitive presence in the Col framework and the modes of engagement in the ICAP framework?

RQ2: If pre-trained language models such as word embeddings are used to automate message labelling, is model performance comparable with prior studies that used linguistically motivated features to train the model?

RQ3: Can an attention layer in a neural network reveal what aspects of a discussion forum message are important for identifying depth and quality of participation?

RQ4: Does model performance improve when labels from Col and ICAP are learned together, compared to the performance of models using each framework separately?

4 METHODOLOGY

My research combines methodological work with quantitative modelling and qualitative content analysis. My current study (target date for completion: early 2020) will compare the Col and ICAP frameworks by looking at co-occurrences of ICAP modes with phases of cognitive presence in a manually labelled data set. This is anonymised data that was collected in a previous study and ethical approval has already been obtained. Specifically, I will approach this task quantitatively by looking at confusion matrices between labels from the two frameworks and visualising them using Epistemic Network Analysis (ENA) (Shaffer, Collier, & Ruis, 2016), as well as comparing the frameworks theoretically and conceptually (**RQ1**). My expectation is that the two frameworks are sufficiently distinct that they will provide complementary insights into the learning processes demonstrated in discussion forum messages.

Later studies will look at automating the labelling of discussion forum messages using advanced NLP methods. I will develop neural network models that incorporate word embeddings and an attention layer (target date for completion: April 2020) and compare the performance of these models with simpler predictive models such as random forests – both quantitatively, in terms of model performance (**RQ2**), and also qualitatively, in terms of potential explanatory power (**RQ3**). By mapping the words into a high-dimensional vector space using word embeddings, the effects of

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

particular word choices are expected to diminish. Therefore, I expect that this approach could prove to be just as powerful as using linguistically motivated model features, while adding flexibility. An attention layer could indicate which words and phrases best characterise the depth and quality of participation according to each of the theoretical frameworks. These results can be validated qualitatively by comparison with prior work on factors contributing to student engagement. One potential future application of this aspect of the research could perhaps be the automatic generation of hints for students about how to improve their own discussion contributions. Finally, I will use multi-task and transfer learning to train models using the labels from both frameworks at once (target date for completion: July 2020) and compare their performance with models trained on each set of labels individually, addressing **RQ4**. If performance improves, in line with prior work, this result would also support the use of both frameworks together in future studies.

5 CURRENT STATUS AND RESULTS ACHIEVED

My first methodological study looking at how data contamination can arise from common data pre-processing practices was presented at LAK'19. A summary of this work was also shared with a broad data science audience at a UK-wide workshop (Advances in Data Science 2019). I am now working on data preparation for my next experimental study. I have adapted the extended ICAP coding scheme used in prior work (Yogev, Gal, Karger, Facciotti, & Igo, 2018) to be more relevant to the context of the data set that I am using. The messages were already labelled with phases of cognitive presence, and manual annotation with the labels from my adapted ICAP scheme is in progress. A study based on preliminary analysis of this data was accepted as a short paper in the main LAK'20 research track.

REFERENCES

- Chi, M. T., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, 49(4), 219-243. doi:https://doi.org/10.1080/00461520.2014.965823
- Collobert, R., & Weston, J. (2008). A unified architecture for natural language processing. *Proceedings of the 25th International Conference on Machine Learning* (pp. 160-167). New York, NY, USA: ACM. doi:https://doi.org/10.1145/1390156.1390177
- Farrow, E. (2018). *Automated content analysis for modelling student engagement in online discussion forums*. Master's Thesis, School of Informatics. University of Edinburgh.
- Farrow, E., Moore, J., & Gašević, D. (2019). Analysing discussion forum data: a replication study avoiding data contamination. *Proceedings of the 9th International Learning Analytics & Knowledge Conference (LAK19)*, (pp. 1-10). doi:https://doi.org/10.1145/3303772.3303779
- Garrison, D. R., Anderson, T., & Archer, W. (1999). Critical Inquiry in a Text-Based Environment: Computer Conferencing in Higher Education. *The Internet and Higher Education*, 2(2), 87-105. doi:https://doi.org/10.1016/S1096-7516(00)00016-6
- Gašević, D., Adesope, O., Joksimović, S., & Kovanović, V. (2015). Externally-facilitated regulation scaffolding and role assignment to develop cognitive presence in asynchronous online discussions. *The Internet and Higher Education*, 24, 53-65. doi:https://doi.org/10.1016/j.iheduc.2014.09.006
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Kovanović, V., Joksimović, S., Waters, Z., Gašević, D., Kitto, K., Hatala, M., & Siemens, G. (2016). Towards Automated Content Analysis of Discussion Transcripts: A Cognitive Presence Case.

- Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 15-24). New York, NY, USA: ACM. doi:<https://doi.org/10.1145/2883851.2883950>
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge: Cambridge University Press. doi:<https://doi.org/10.1017/CBO9780511894664>
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2* (pp. 3111-3119). Curran Associates, Inc.
- Neto, V., Rolim, V., Ferreira, R., Kovanović, V., Gašević, D., Dueire Lins, R., & Lins, R. (2018). Automated Analysis of Cognitive Presence in Online Discussions Written in Portuguese. *Proceedings of the 13th European Conference on Technology Enhanced Learning* (pp. 245-261). Cham: Springer International Publishing. doi:https://doi.org/10.1007/978-3-319-98572-5_19
- Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global Vectors for Word Representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532-1543). Association for Computational Linguistics. doi:<https://doi.org/10.3115/v1/D14-1162>
- Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A Tutorial on Epistemic Network Analysis: Analyzing the Structure of Connections in Cognitive, Social, and Interaction Data. *Journal of Learning Analytics*, 3(3), 9-45. doi:<https://doi.org/10.18608/jla.2016.33.3>
- Taskin, Y., Hecking, T., Hoppe, H. U., Dimitrova, V., & Mitrovic, A. (2019). Characterizing Comment Types and Levels of Engagement in Video-Based Learning as a Basis for Adaptive Nudging. *Proceedings of the 14th European Conference on Technology Enhanced Learning* (pp. 362-376). Cham: Springer International Publishing.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, 29(1), 24-54. doi:<https://doi.org/10.1177/0261927X09351676>
- Wang, X., Wen, M., & Rosé, C. P. (2016). Towards Triggering Higher-order Thinking Behaviors in MOOCs. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 398-407). New York, NY, USA: ACM. doi:<https://doi.org/10.1145/2883851.2883964>
- Wang, X., Yang, D., Wen, M., Koedinger, K., & Rosé, C. P. (2015). Investigating how student's cognitive behavior in MOOC discussion forums affect learning gains. *Proceedings of the 8th International Conference on Educational Data Mining* (pp. 226-233). International Educational Data Mining Society (IEDMS).
- Waters, Z., Kovanović, V., Kitto, K., & Gašević, D. (2015). Structure Matters: Adoption of Structured Classification Approach in the Context of Cognitive Presence Classification. *Information Retrieval Technology*, 9460, 227-238. doi:https://doi.org/10.1007/978-3-319-28940-3_18
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., . . . Dean, J. (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. *CoRR*. Retrieved from <http://arxiv.org/abs/1609.08144>
- Yogev, E., Gal, K., Karger, D., Facciotti, M. T., & Igo, M. (2018). Classifying and Visualizing Students' Cognitive Engagement in Course Readings. *Proceedings of the Fifth Annual ACM Conference on Learning at Scale* (pp. 1-10). New York, NY, USA: ACM. doi:10.1145/3231644.3231648

Guidance in Multimodal Learning Analytics for Collaborative Classrooms

Gloria Milena Fernandez-Nieto
University of Technology Sydney
gloria.m.fernandeznieto@student.uts.edu.au

ABSTRACT. Feedback is a crucial aspect of classroom-based learning. Delivering high-quality feedback can help students to make well-informed decisions by understanding their learning goals and current performance in light of the teacher's pedagogical intentions. However, providing actionable feedback in the physical classroom can be challenging for teachers, especially in large classes with many students or groups to track. One way to provide such feedback automatically is through learning analytics (LA) visual interfaces, in which digital traces and analytics outputs are shown to teachers and students. However, recent research highlights concerns about the complexity of LA interfaces, lack of guidance for students to gain insights, and lack of educationally meaningful impact. Moreover, the technical challenge of capturing, displaying, and making sense of data about collocated activity in the classroom is more complex than in fully computer-mediated settings. In response, this doctoral thesis aims to address the following research question: *how to communicate educationally meaningful feedback from multimodal data collected from collocated collaboration to provoke reflection on teaching and learning?* This paper presents a brief motivation for this thesis, the methodology to be followed and the current state of the project.

Keywords: feedback, visualisation, guidance, reflection, collocated spaces, multimodal data

1 MOTIVATION AND PROBLEM DESCRIPTION

In the 21st century classroom, students collaborate and communicate both online and physically. Collaboration, team work, communication, problem solving, among other skills were labelled as *21st century skills* to indicate their relevance in the new century, and claims abound that people need to acquire such skills to improve their professional proficiency (Voogt, Erstad, Dede, & Mishra, 2013). Collaboration skills allow students to solve problems, engage in inquiry-based activities (e.g. science experiments) or research particular topics. The provision of high-quality actionable feedback, can have a strong effect on student reflection, performance and achievement (Hattie & Timperley, 2007) (Figure 1 Thesis diagram overview, a). Actionable feedback, can help students to take actions to close the gap between where they are and where they are aiming to be. However, providing actionable feedback is very challenging especially for large classes (Winstone & Carless, 2019) and in physical spaces, where a teacher needs to follow the progress of each student, or group of students, who are often working concurrently.

One promising strategy to collect evidence from physical spaces involves the use of emerging sensing technology and interactive devices (e.g. Martinez-Maldonado et al., 2016). Data of an specific phenomenon collected from different types of devices, sensors, measurement techniques, experimental setups and other sources is called *multimodal data*. *Multimodal Learning Analytics* (MMLA) is an emerging sub-area within the broader field of Learning Analytics, which focuses on the use of multimodal data with educational theories, to understand and support learning and teaching

processes (Schneider, Di Mitri, Limbu, & Drachsler, 2018). MMLA applications, can potentially be used to support teachers delivering automatic or semiautomatic feedback in the classroom because it can offer new insights about situations that have been for a long time considered ephemeral, such as collaboration or high effective collocated teamwork in the workplace. In the Collaborative Classroom (CC), multimodal data introduces new opportunities for generating a better understanding about teaching and learning processes in the classroom to enhance the provision of feedback (Järvelä, Malmberg, Haataja, Sobocinski, & Kirschner, 2019) (see Figure 1, b, opportunity 1). Yet, a critical challenge is how to use these rich data to create effective MMLA feedback interfaces.

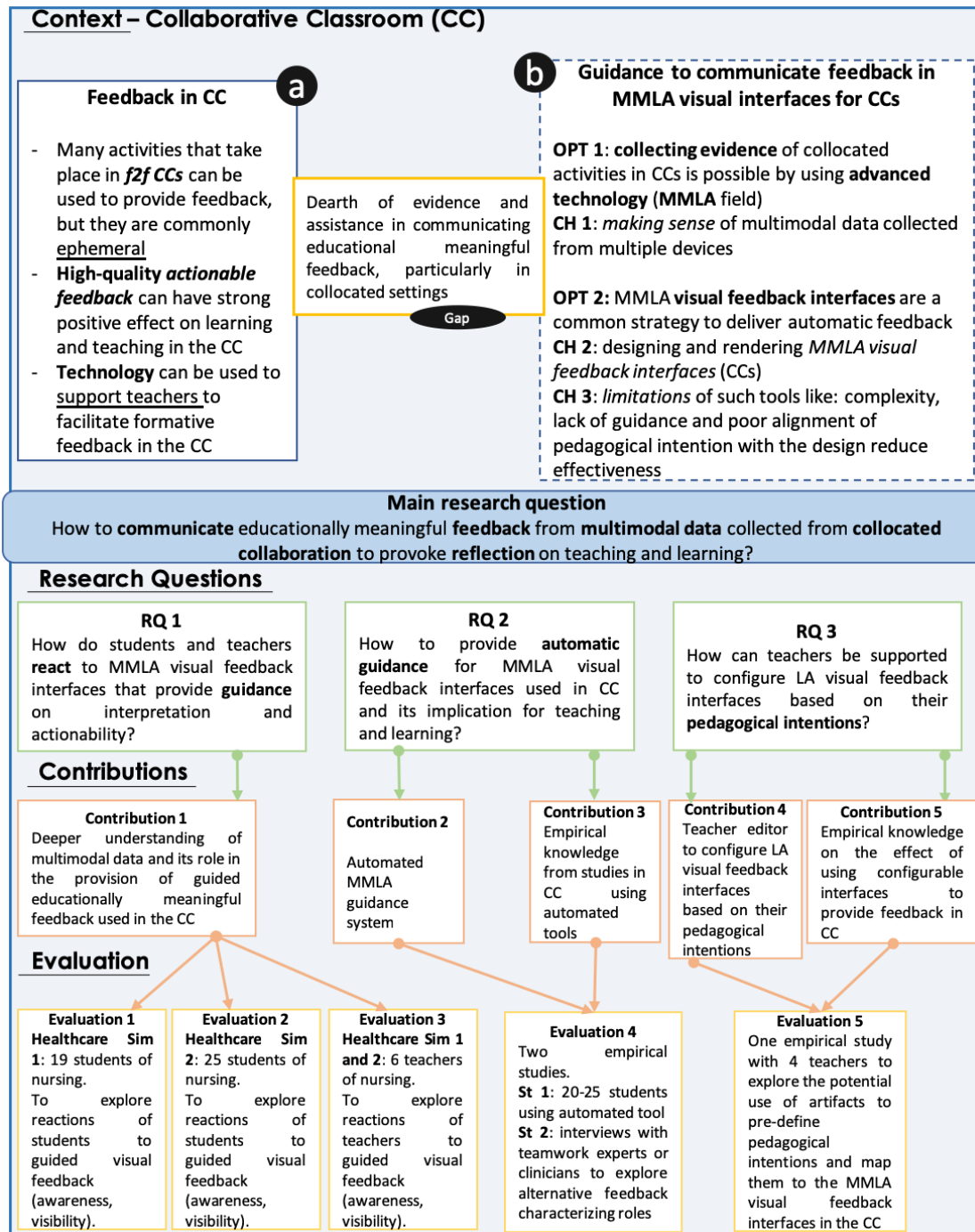


Figure 1 Thesis diagram overview

Researchers have explored benefits of visual feedback in (i) helping educators and students to gain a better understanding of their teaching and learning processes (e.g. Klein et al.); (ii) prompting reflections (e.g. Bodily & Verbert, 2017); and (iii) making informed decisions about teaching and learning (e.g. Jivet, Scheffel, Drachsler, & Specht, 2017). The MMLA community has a small but growing interest in creating MMLA visual feedback interfaces to visually communicate feedback to students and teachers (e.g. Domínguez, Chiluiza, Echeverria, & Ochoa, 2015) (see Figure 1, b, opportunity 2).

However, capturing, rendering visible and making sense of multimodal data collected from physical spaces to deliver *actionable feedback* to *frequent* users, who bring little or no data analysis expertise (Schield, 2004) or *casual* users, who use a system occasionally (Cuff, 1980), is an actual challenge. This challenge may not only involve technical aspects (e.g. interoperability), but also human factors (how can people interpret and interact with data?), pedagogical aspects (how can teachers appropriate data-intensive solutions into their practice?), learning aspects (how can learners reflect on evidence?) and the characteristics of the learning design (what is the meaning of data in a particular context?) (Figure 1 Thesis diagram overview, b, challenges 1). Although the MMLA community is acknowledging some of these challenges, a dearth of research has sought to advance in the design of effective MMLA visual feedback interfaces (Figure 1, b, challenges 2). Specifically, there is a lack of tailored visual interfaces to assist students and teachers (casual users) to interpret and make effective, pedagogical use of data in the context of colocated teamwork and group activity.

A range of literature demonstrates that it is naïve to expect that students and teachers will be able to interpret data and make them actionable without further guidance and support (Bodily & Verbert, 2017) (Figure 1, b, challenges 3). This thesis argues that providing assistance and additional guidance can support teachers and students in understanding educationally meaningful feedback on colocated activities in the CC. *Guidance*, can be defined as a computer-assisted processes aimed at narrowing the gap of data interpretation and exploration encountered by end-users (Ceneda et al., 2017). Schultz et al. (Schulz, Streit, May, & Tominski, 2013) described different ways in which this concept can be materialised, such as by enhancing charts using visual cues, allowing users to select from various visualisation techniques, and guiding users through prescriptive data exploration workflows or via *data storytelling*. The Data Storytelling approach, which is a suite of information design and ‘compression’ techniques to help an audience effectively understand what is important in a visualisation (Ryan, 2016), is one alternative to provide additional guidance in MMLA visual feedback interfaces.

In short, making sense of multimodal data collected from CCs to provide educationally meaningful feedback is difficult. Coupled with the challenges of interpreting visual feedback tools, this problem makes it difficult to provide actionable feedback for students and teachers that could help them reflect on teaching and learning. This concern becomes even more profound for the case of MMLA visual feedback interfaces, which commonly deal with complex heterogeneous data streams (see Figure 1 Thesis diagram overview, gap). This thesis addresses the dearth of evidence and assistance in communicating educational meaningful feedback collected from authentic CCs. This research aims (i) to investigate students’ and teachers’ reactions to MMLA visual feedback interfaces that provides guidance in interpretation and actionability (Figure 1, RQ1), (ii) to provide automatic guidance for

MMLA visual feedback interfaces (Figure 1 Thesis diagram overview, RQ2), and (iii) to support teachers to configure LA visual feedback interfaces based on the pedagogical intentions of the CC (Figure 1, RQ3).

2 METHODOLOGY

The selected methodology to tackle the research aims of this thesis is Mixed Methods Research. This methodology is commonly used for research processes to integrate quantitative and qualitative methods of data collection and analysis (Plano Clark & Ivankova, 2016). As described in (Ivankova & Stick, 2007) the designing sequence of a mixed method research includes the following phases, (a) quantitative data collection, (b) quantitative data analysis, (c) case selection; interview development protocol, (d) qualitative data collection, (e) qualitative data analysis, and (f) integration of quantitative and qualitative results. The procedures and products of each phase defined for this doctoral thesis are explain as follows.

The quantitative data (a), will be captured from authentic collaborative classrooms (e.g. health simulations, physics labs etc.) by using Empatica e4 bracelets (to captured physiological data from students and teachers), actions performed by students (e.g. stopping intravenous-IV fluid, writing on charts and calling the doctor) manually logged by an observer, positioning wearable tags (Pozyx.io), audio recorders and video. Analysis of the data collected (b) will vary depending on the device (e.g. peaks of stress can be extracted from the electrodermal activity -EDA data captured from the bracelets). The objective of this phase is to translate raw data into educationally meaningful insights by using artifacts like the *Multimodal Matrix (MM)* described by Echeverria et al. (2019). Once the data is analysed the MMLA visual interface will be generated and enhanced by using visualisation guidance approaches to easily communicate insights.

The qualitative cases that are selected (c) will use a retrospective reflection technique (Hassenzahl & Ullrich, 2007) to investigate the opportunities and challenges of the LA visual interfaces prototypes to communicate insights and its impact on learning and teaching processes. The qualitative data collection (d) will consist on interviews using LATEP (Learning Analytics Translucence Elicitation Process), an elicitation protocol for understanding how non-data experts envisage the use of LA systems (Martinez-Maldonado et al., 2019). Reflection sessions will be audio-recorded, fully transcribed, and coded using NVivo. Following best practices of qualitative research analysis (e) and given the direct alignment between the study protocol and the analysis themes, statements of interest will be coded by two researchers according to the pre-set themes of the study protocol. Finally, to integrate the qualitative and quantitative methods (f) interpretations and explanations of the results will be presented as discussions, implications and future research.

3 CURRENT WORK

This section describes the work that has been done during the first year of the project. The literature review evidenced three different gaps: (i) teachers and students commonly have little evidence to reflect on teaching and learning (addressed by RQ1); (ii) most LA visual interfaces have limitations in terms of complexity, interpretation, lack of guidance, and impact on learning (RQ2); and (iii) MMLA

data multiplies the challenge of interpretation (align with pedagogical intentions) and complexity to deliver visual interfaces (*RQ3*).

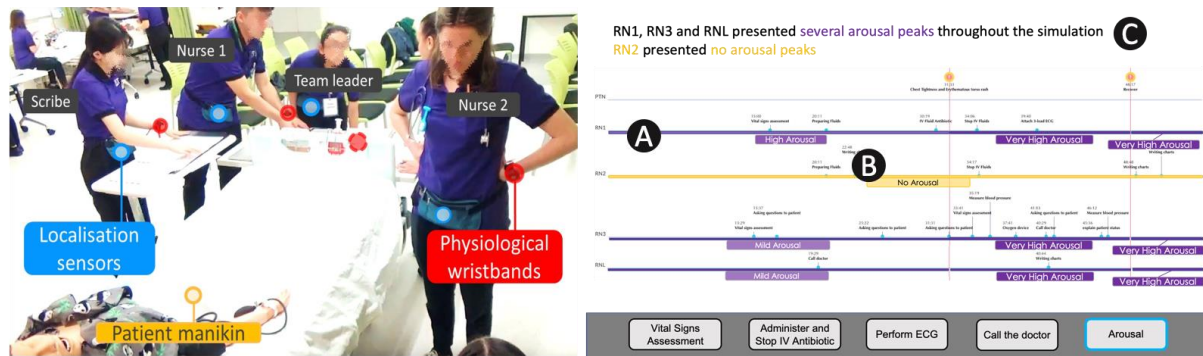


Figure 2. Left: Nursing Simulation. Right: MMLA visual interface, prototype 2 (arousal peaks)

Additionally, two empirical studies were conducted to address *RQ1* and to validate the viability and the scope of this research project. For this purpose, two MMLA visual interfaces prototypes (see an example of prototype 2 in Figure 2, right) were semi-automatically generated (with data collected from actions performed by students and EDA). One week later, the prototypes were presented during focus group interviews to understand the impact and opportunities of the prototypes to communicate insights to students. The approach used to help students effectively understand insights was DS. Figure 2, presents a data story about arousal peaks extracted from the EDA data captured during the simulation. The visualisation uses DS principles (e.g. a data story should guide attention) to communicate key messages through adding enhancements such; enclosing areas (A); changing colour, contrast or thickness; annotating salient data points (B) or adding titles that summarise the take-away message (C).

According to the studies, results showed that the visual enhancements of the visual interfaces, such as text annotations and selectively emphasising parts of a chart, helped students identify misconceptions, think about strategies to address errors they made, and reflect on the arousal peaks they may have experienced during the simulations. In the short term, we plan to conduct studies where an automated or semi-automated version of the prototypes will be included into the learning designs and aligned with pedagogical intentions.

4 CONCLUSIONS

This thesis will explore the potential, impact of guidance in MMLA visual feedback interfaces in mix-method studies conducted in authentic classrooms with the purpose of communicating insights to students and teachers through data stories. Many questions have arisen about ethical implications of using multimodal data, transparency of algorithms, risk of overinterpretation of insights, automatic generation of feedback and scalability of MMLA innovations. Given the previous challenges, the interest of the LA community to help students and teachers make the most of the new forms of feedback that are becoming possible will possibly grow.

REFERENCES

- Bodily, R., & Verbert, K. (2017). Trends and issues in student-facing learning analytics reporting systems research. *7th LAK Conference*, 309–318. New York, USA: ACM Press.
- Ceneda, D., Gschwandtner, T., May, T., Miksch, S., Schulz, H. J., Streit, M., & Tominski, C. (2017). Characterizing Guidance in Visual Analytics. *IEEE Transactions on Visualization and Computer Graphics*, 23(1), 111–120.
- Cuff, R. N. (1980). On casual users. *International Journal of Man-Machine Studies*, 12(2), 163–157.
- Domínguez, F., Chiluíza, K., Echeverría, V., & Ochoa, X. (2015). Multimodal Selfies: Designing a Multimodal Recording Device for Students in Traditional Classrooms. *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction - ICMI '15*, 567–574.
- Hassenzahl, M., & Ullrich, D. (2007). To do or not to do: Differences in user experience and retrospective judgments depending on the presence or absence of instrumental goals. *Interacting with Computers*, 19(4), 429–437.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112.
- Ivankova, N. V., & Stick, S. L. (2007). Students' Persistence in a Distributed Doctoral Program in Educational Leadership in Higher Education: A Mixed Methods Study. *Research in Higher Education*, 48(1), 93–135. <https://doi.org/10.1007/s11162-006-9025-4>
- Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. A. (2019). What multimodal data can tell us about the students' regulation of their learning process? *Learning and Instruction*, (June), 0–1. <https://doi.org/10.1016/j.learninstruc.2019.04.004>
- Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness Is Not Enough: Pitfalls of Learning Analytics Dashboards in the Educational Practice. In *European Conference on Technology Enhanced Learning* (pp. 82–96). Tallinn, Estonia: Springer, Cham.
- Klein, C., Lester, J., Nguyen, T., Justen, A., Rangwala, H., & Johri, A. (2019). Student Sensemaking of Learning Analytics Dashboard Interventions in Higher Education. *Journal of Educational Technology Systems*, 48(1), 130–154. <https://doi.org/10.1177/0047239519859854>
- Martinez-Maldonado, R., Echeverría, V., Elliott, D., Axisa, C., Power, T., & Buckingham Shum, S. (2019). Designing translucent learning analytics with teachers: an elicitation process. *Interactive Learning Environments*.
- Martinez-Maldonado, R., Schneider, B., Charleer, S., Shum, S. B., Klerkx, J., & Duval, E. (2016). Interactive surfaces and learning analytics: data, orchestration aspects, pedagogical uses and challenges. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*, 124–133. New York, New York, USA: ACM Press.
- Plano Clark, V. L., & Ivankova, N. V. (2016). *Mixed Methods Research: A Guide to the Field*. <https://doi.org/10.4135/9781483398341>
- Ryan, L. (2016). *The visual imperative : creating a visual culture of data discovery*. Massachusetts: Kaufmann, Morgan.
- Schild, M. (2004). Information Literacy, Statistical Literacy and Data Literacy. *IASSIST*, 28, 7–14.
- Schneider, J., Di Mitri, D., Limbu, B., & Drachsler, H. (2018). *Multimodal Learning Hub: A Tool for Capturing Customizable Multimodal Learning Experiences*. https://doi.org/10.1007/978-3-319-98572-5_4
- Schulz, H.-J., Streit, M., May, T., & Tominski, C. (2013). Towards a Characterization of Guidance in Visualization. In *Poster at the IEEE InfoVis'13*.
- Voogt, J., Erstad, O., Dede, C., & Mishra, P. (2013). Challenges to learning and schooling in the digital networked world of the 21st century. *Journal of Computer Assisted Learning*, 29(5), 403–413.
- Vytasek, J. M., Patzak, A., & Winne, P. H. (2020). *Analytics for Student Engagement*. https://doi.org/10.1007/978-3-030-13743-4_3
- Winstone, N. E., & Carless, D. (2019). *Designing effective feedback processes in higher education : a learning-focused approach* (1st ed.; S. for research into higher education Series, Ed.). Research into Higher Education.

Learning Analytics and Capability Approach in Education: Analysing Student Agency in Higher Education

Ville Heilala

University of Jyväskylä, Finland

ville.s.heilala@jyu.fi

ABSTRACT: An important direction of learning analytics research is combining learning analytics with the theoretical foundations of education. In the quest to find some solid basis for my research in the field, I present some of my initial ideas of combining learning analytics with the capability approach in education. In this context, agency is an essential concept when considering the broader use of the capability approach. After briefly introducing relevant concepts, I present our learning analytics research done relating to student agency, including future research ideas. My research contributes to the discussion about the concept of student agency, and research on how learning analytics could be used to inform pedagogical and learning practices from the both student and teacher point of view.

Keywords: capability approach, student agency, learning analytics, robust clustering

1 INTRODUCTION

Learning analytics is based on the measurement, collection, analysis, reporting, and interpreting of learner-generated data in different learning contexts (Conole, 2011). Recent literature reviews on learning analytics have identified several issues, which might be essential to be addressed in future research in the field. Banihashem et al. (2018, p. 8) have reviewed learning analytics from the educational point of view, and they conclude that “it is a matter of emergency to deeply understand what learning analytics is and how it could be widely used in the educational settings.” They emphasize the understanding of learning and teaching, and state that innovations and technologies in the field “should be passed through the filter of the theoretical foundations of education” (ibid. p. 7). The previous arguments are currently crucial because there is still very little evidence about the effectiveness of learning analytics interventions (Sønderlund, 2018).

When considering other fields that are encompassed by data-driven approach, for example, business and healthcare, we can identify aims like increasing profitability, improving customer engagement, operations efficiency, predictive risk management, and clinical decision support (e.g., Schniederjans et al., 2014; Simpao et al., 2014). The analytics and methods involved in the cases mentioned above are grounded in the overall aims. Furthermore, the aim dictates what kinds of data are needed and what kinds of measures should be used. Collecting the data is also one of the first steps in both conducting learning analytics research and applying it in practice. Thus, it is essential to consider the starting point: what are the educational aims we are trying to achieve when developing and applying learning analytics? Biesta (2010, p. 26) argues that currently, a discussion about education is “dominated by the measurement of educational outcomes” and the danger “is that we end up valuing what is measured, rather than that we engage in measurement of what we value.” By using the Foucault’s (1975) notion about benthamian panopticon, Wintrup (2017) depicts the situation in

which learning analytics might introduce unanticipated consequences: measuring, observing, and supervising students may lead to the students behaving thought they are “being observed and changing behaviors accordingly” (ibid., p. 39). Also, Wise (2013, p. 52) points out the concern that when utilizing learning analytics, “the analytics alone will dictate how people engage in the learning activity.” As one possible solution, she suggests both using multiple diverse measures in analyzing learning and supporting students in interpreting the meaning of the analytics results.

Before we can do meaningful learning analytics, we need to address the question of *why* before *what* and *how*. In trying to ground my research on learning analytics to an educationally sound basis, I resort to the discussion about the aims of education. Further, I explore what Amartya Sen’s capability approach might provide for learning analytics research. Lastly, I present some results of our ongoing work done on student agency analytics.

1.1 The aims of good education

The aims of education have been under a philosophical discussion for centuries. At this point, I refer briefly to Biesta (2010), who presents three functions of education: qualification, socialization, and subjectification. Qualification refers to an educational function of providing learners knowledge, skills, and understanding so they can accomplish their endeavors. Socialization refers to the ways how education enables us to become part of the social, cultural, and political whole. Subjectification is the process opposite to socialization: learners are also individuals and separate from the encompassing order. The aims of good education provide us some guidelines of what we need to learn and teach. What we need to learn and teach gives us some hints about what we need to measure and evaluate. However, as Biesta (2010, p. 128) argues, “the question of good education is a normative question,” and it cannot “be answered by the outcomes of measurement, by research evidence or through managerial forms of accountability.” Fortunately, he concludes that “education should always entail an orientation toward freedom” (Biesta, 2010, p. 129). If education has an orientation toward freedom, what kind of learning analytics we should engage in? For me, the question has led to the discovery of Amartya Sen’s capability approach.

1.2 Capability approach in education

The capability approach is a framework for development, a good life, and well-being, and many scholars have developed it (e.g., Martha Nussbaum). However, at this point, I concentrate on the views of a Nobel-prize winner economist and a philosopher Amartya Sen. The core ideas of the capability approach consist of functionings and capabilities. According to Sen (1987, p. 36), “a functioning is an achievement, whereas a capability is the ability to achieve.” In other words, functionings refers to “various states of human beings and activities that a person can undertake” and capabilities are “a person’s real freedoms or opportunities to achieve functionings” (Robeyns, 2011). Most notably, agency is a central concept in the realization of the capabilities, functionings, and freedom (Crocker, 2008). Saito (2003) argues that education might play a role in enhancing capacities and opportunities. Thus, we might pose a question: can enhancing agency somehow contribute to the realization of the capabilities, functionings, and freedom in education? In the next section, I introduce the concept of agency in the context of education.

1.3 Student agency

The discussion about human agency has its roots, for example, in social theory (e.g., Giddens, 1984) and social cognitive theory (Bandura, 2001). One of the most thorough examinations of agency has been presented by Emirbayer and Mische (1998). In the context of education, student agency has been brought recently into the discussion at the educational policy level. In the OECD Future of Education and Skills 2030 framework, student agency is defined as follows (OECD, 2019, p. 4):

Student agency relates to the development of an **identity** and a **sense of belonging**. When students develop agency they rely on **motivation, hope, self-efficacy** and a **growth mindset** (the understanding that abilities and intelligence can be developed) to navigate towards well-being. This enables them to act with a sense of **purpose**, which guides them to flourish and thrive in society.

However, the previous research relating to student agency or its operationalization is scarce. In our research group, we utilize the validated Agency of University Students (AUS) scale developed by Jääskelä et al. (2016) and the related conceptualization, which defines student agency in higher education as “access to (and use of) resources for purposeful action in study contexts, that is, as students’ experienced or interpreted individual, interactional and contextual resources to engage in intentional and meaningful action and learning” (ibid., p. 7). The scale measures student agency in 11 dimensions of agency: competence beliefs, self-efficacy, interest and utility value, participation activity, ease of participation, opportunities to influence, opportunities to make choices, peer support, equal treatment, trust, and teacher support.

1.4 Research questions

The general aims of my current dissertation research are 1) to establish a strong foundation for my research based on educational theories and 2) examine the effectiveness of the student agency analytics in higher education. Thus, I set the following research questions:

RQ1: How capability approach could contribute to the learning analytics research?

RQ2: Is it possible to use the student agency analytics process to effectively inform pedagogical practices?

2 STUDENT AGENCY ANALYTICS

I intend to answer the research questions mentioned above using a process called student agency analytics (Jääskelä, Heilala et al., in revision). The process is based on the validated AUS scale, robust clustering, and service-based automatic provisioning of the analytics at scale. The purpose of student agency analytics is to provide personal feedback to the student and summarized agency information for the teacher. The information could be used, for example, in self-reflection and pedagogical decision making. In the following sections, I briefly present the analytics process and our recent research findings.

2.1 Method

From the learning analytics point of view, we utilize robust unsupervised clustering methods (Kärkkäinen & Heikkola, 2004; Hämmäläinen, 2018) to discover novel information relating to student agency. The proposed student agency analytics process consists: 1) acquiring the agency data using the AUS questionnaire, 2) preprocessing the data and imputing the missing values, 3) calculating the individual agency factors using the factor pattern matrix, 4) derive agency profiles using robust clustering, and 5) presenting individual results to students and aggregated information to teachers.

It is essential to take into account the ethical requirements when conducting learning analytics. From the technological point of view, the student agency analytics process will utilize a service-based architecture (Figure 1). Concerning the General Data Protection Regulation, the data between data controller (e.g., educational institution) and data processor (e.g., the learning analytics service provider) is transmitted as pseudonymized. The architectural approach utilizing pseudonymization of the data takes into account some of the legal and ethical requirements of learning analytics.

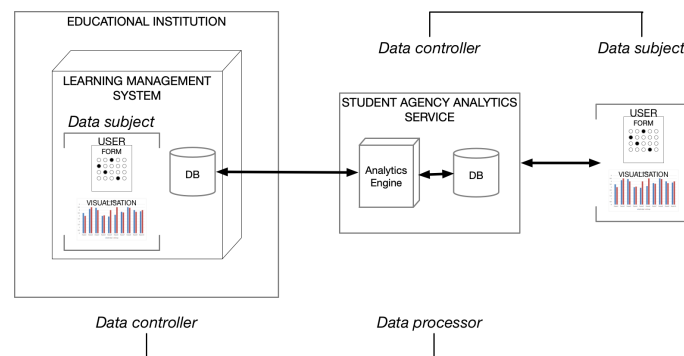


Figure 1. Service-based architecture of student agency analytics service.

2.2 The results achieved so far

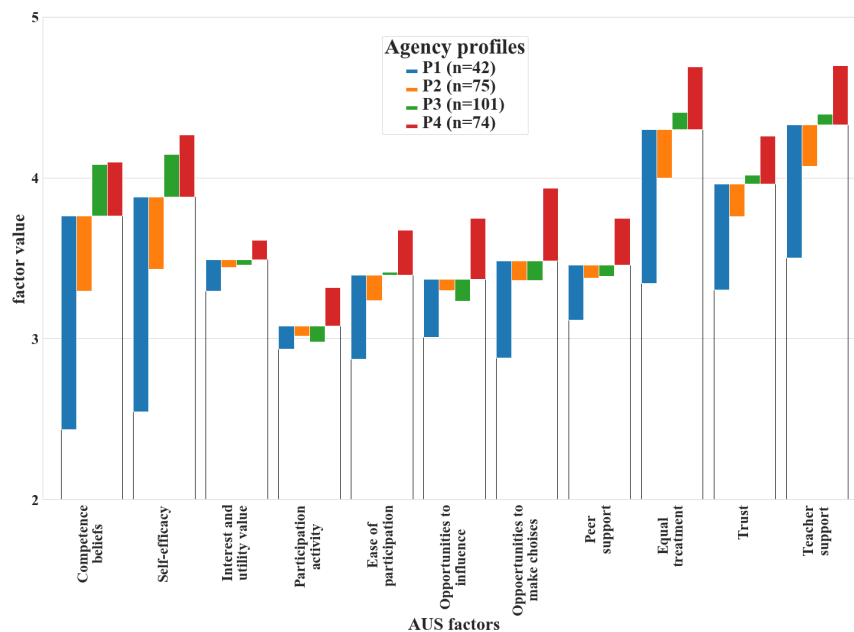


Figure 2. General student agency level and the deviations of the different agency profiles.

The agency analytics process provides information about four different student agency profiles among students in a particular learning context. Figure 1 presents an example of the agency analytics results in a university course. For example, the agency profile P1 depicts the low agency profile. The students in P1 assessed their competence beliefs, self-efficacy, and teacher-student relational resources of agency (e.g., teacher support) as lower than students in other profiles. Relating to the AUS questionnaire, students can provide open-ended answers about the restrictive and supporting aspects of the course. By analyzing the open-ended answers, it is possible to acquire more detailed information about the students' study experiences in different agency profiles (Heilala et al., forthcoming). It might also be possible to connect student agency to course outcomes (Jääskelä, Heilala et al., in revision).

3 DISCUSSION

In this paper, I presented my preliminary and tentative thoughts about combining Sen's capability approach with learning analytics using student agency analytics as an example. I pose one definition of learning analytics, which I hope to help focus my research: *Learning analytics is the use of data to enhance the capabilities and the functionings of an entity in relation to its purpose in the system of education*. An entity, in this case, can be a learner, a teacher, a study advisor, an educational organization, or anything or anyone relating to purposeful educational activity or learning.

Student agency analytics was presented as one possible application of the learning analytics definition mentioned above. Wise (2014) emphasizes the importance of agency in learning by setting it as one of the four principles of pedagogical learning analytics intervention design. Similarly, the purpose of our student agency analytics is to provide meaningful information for the students and teachers for the basis of self-reflection and pedagogical decision making. The research continues the discussion concerning the construct of student agency by emphasizing the multidimensional nature of agency. Furthermore, the research contributes to the research on how learning analytics could be used to inform pedagogical and learning practices from the both student and teacher point of view.

However, further research is needed to find evidence about the effectiveness of the agency analytics process. My future research might involve, for example, examining the agency analytics process in learning analytics interventions (e.g., Could student agency analytics be used to inform pedagogical practices effectively?), further development of the AUS Scale, and methodological development (e.g., analysis of open-ended textual course feedback using NLP methods).

REFERENCES

- Bandura, A. (2001). Social cognitive theory: an agentic perspective. *Annual Review of Psychology*, 52.
- Banihashem, S. K., Aliabadi, K., Ardakani, S., Delaver, A., & Ahmadabadi, M. (2018). Learning analytics: a systematic literature review. *Interdisciplinary Journal of Virtual Learning in Medical Sciences*, 9(2).
- Biesta, G. J. J. (2010). *Good education in an age of measurement: Ethics, politics, democracy*. Boulder, CO: Paradigm Publishers.
- Conole, G., Gašević, D., Long, P., & Siemens, G. (2011). Message from the LAK 2011 General & Program Chairs. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*. New York, NY: ACM.

- Crocker, D. A. (2008). *Ethics of global development: Agency, capability, and deliberative democracy*. Cambridge, England: Cambridge University Press.
- Emirbayer, M., & Mische, A. (1998). What is agency? *American Journal of Sociology*, 103(4), 962–1023.
- Foucault, M. (1975). *Discipline and punish: The birth of the prison*. London: Penguin.
- Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. Cambridge, England: Polity Press.
- Heilala, V., Jääskelä, P., Kärkkäinen, T., & Saarela, M. (forthcoming). Understanding the Study Experiences of Students in Low Agency Profile: Towards a Smart Education Approach. *Proceedings of the First International Conference on Smart & Communication Technologies SmartICT'19*. In: Leopold Angrisani et al. *Lecture notes in electrical engineering*. Berlin: Springer.
- Hämäläinen, J. (2018). *Improvements and applications of the elements of prototype-based clustering* (Doctoral dissertation, University of Jyväskylä, Finland). Retrieved from <http://urn.fi/URN:ISBN:978-951-39-7621-7>
- Jääskelä, P., Heilala, V., Häkkinen, P., & Kärkkäinen, T. (in revision). Student agency analytics: Learning analytics as a tool for analyzing student agency in higher education.
- Jääskelä, P., Poikkeus, A.-M., Vasalampi, K., Valleala, U. M., & Rasku-Puttonen, H. (2016). Assessing agency of university students: validation of the AUS Scale. *Studies in Higher Education* 42(11), 1–19.
- Kärkkäinen, T., Heikkola, E. (2004). Robust formulations for training multilayer perceptrons. *Neural Computation* 16(4), 837–862.
- OECD. (2019). *Concept note: student agency for 2030*. Retrieved from https://www.oecd.org/education/2030-project/teaching-and-learning/learning/student-agency/Student_Agency_for_2030_concept_note.pdf
- Robeyns, I. (2011). The capability approach. *Stanford Encyclopedia of Philosophy*. Retrieved from <https://plato.stanford.edu/entries/capability-approach/>
- Schniederjans, M. J., Schniederjans, D. G., & Starkey, C. M. (2014). *Business analytics principles, concepts, and applications: what, why, and how*. London: Pearson Education.
- Simpao, A., Ahumada, A. Gálvez, J., & Rehman, M. (2014). A review of analytics and clinical informatics in health care. *Journal of Medical Systems*, 38(45), 38–45.
- Sønderlund, A., Hughes, E., & Smith, J. (2018). The efficacy of learning analytics interventions in higher education: A systematic review. *British Journal of Educational Technology*, 31(209), 2594–2618.
- Wintrup, J. (2017). Higher education's panopticon? Learning analytics, ethics and student engagement. *Higher Education Policy*, 30(1), 87–103.
- Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, 203–211. New York, NY: ACM.
- Wise, A. F., Zhao, Y., & Hausknecht, S. N. (2013). Learning analytics for online discussions: A pedagogical model for Intervention with embedded and extracted analytics. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 48–56. New York, NY: ACM.

Discovering Teachers' Regulatory Learning Processes in Technology Integration Using Educational Data Mining Approaches

Lingyun Huang

Department of Educational and Counselling Psychology, McGill University
lingyun.huang@mail.mcgill.ca

Susanne P. Lajoie

Department of Educational and Counselling Psychology, McGill University
susanne.lajoie@mcgill.ca

ABSTRACT: Self-regulated learning (SRL) is critical for teachers to gain a sophisticated understanding of technological pedagogical content knowledge (TPACK), which is needed to optimize the use of technologies in teaching. This research aims to discover teachers' SRL processes from 70 participants. We apply educational data mining and learning analytics methods to mine teachers' SRL processes using the computer logs extracted from nBrowser—a computer-based learning environment. The fuzzy mining algorithm of process mining is used to discover the temporal SRL process models. The conformance checking algorithm tests whether there is a significant difference in SRL patterns between teachers with different TPACK performance. The findings can contribute to the advancement of our scientific understanding of the role of SRL in teacher education and inform teacher educators and researchers about how to design scaffolds to support teachers' regulation in complex technology-integrating tasks.

Keywords: Self-regulated learning; Traces Methods; Educational data mining; Process mining

1 INTRODUCTION

Teachers' use of technology to optimize teaching and learning depends on their understanding of the relationship between the subject matter, pedagogical strategies, technological functions, as well as demands of students and contexts. This understanding is referred to as technological pedagogical content knowledge (TPACK, Angeli & Valanides, 2009; Mishra & Koehler, 2006). A conceptual understanding of TPACK allows teachers to monitor and control their ways of implementing technologies and use them in an adaptive fashion. In contrast, a shallow understanding may lead to ineffective technology use, which might distract students from deep learning. Research suggests that computer-based learning environments (CBLEs) are effective for developing conceptual understandings of complex knowledge (Azevedo, Guthrie, & Seibert, 2004; Greene, Bolick, & Robertson, 2010). It is in part due to the fact that CBLEs can be used as cognitive and metacognitive tools that foster students' self-regulation of learning (SRL) that demonstrates a significant impact on deeply learning complex knowledge (Azevedo, 2005; Lajoie, 2000). The information-processing perspective of self-regulated learning (IP-SRL, Winne & Hadwin, 1998) model accounts for the cognitive and metacognitive processes effective for conceptual understandings of complex knowledge. Such processes include analyzing tasks, goal setting and strategy planning, monitoring and adapting. These processes are interrelated and the consequences of the previous process influence the subsequent ones (Winne, 2011). Accordingly, learners are supposed to monitor the

process to ensure the quality of learning. In this work, we employ the IP-SRL model to conceptualize teachers' TPACK acquisition. Teachers first need to understand the task, retrieve prior knowledge, analyze the characteristics of the subject contents, and evaluate motivation or beliefs. Next, teachers set instructional goals and plan specific pedagogical, tools, and classroom management strategies. Then teachers design lessons by incorporating relevant learning materials and tools and coordinating planned strategies. Finally, teachers reflect their design based on the framework of TPACK and adjust the ways of using technologies accordingly. The initial three steps justify how teacher coordinate the affordances of technologies with subject contents and pedagogical strategies to select the most appropriate tools. The final step justifies how teachers use TPACK as metaknowledge to evaluate the overall performance of technology use. Meanwhile, self-regulated teachers also use TPACK to metacognitively monitor and evaluate each subphase along with the whole process to achieve their goals and make an adaptation whenever it is necessary. Based on such, we assume that teachers who are self-regulated will attain deep levels of TPACK understanding

Several studies examine the influence of teachers' self-regulation in TPACK development (Kramarski & Michalsky, 2010; Poitras, Mayne, Huang, Udy, & Lajoie, 2018). However, there are still spaces for further discussions. More concretely, the recent conceptualization of SRL as a sequence of events (Winne & Perry, 2000) and microanalysis of SRL processes (Azevedo et al., 2004) allows for exploring how teachers regulate their TPACK learning and which specific SRL processes contribute to deep learning. In light of methods, more fine-grained data extracted from CBLEs and the advances in educational data mining and learning analytics provide the potential to model and scrutinize teachers' self-regulatory processes (Bannert, Molenaar, Azevedo, Järvelä, & Gašević, 2017). Traces or trace-logs, for example, are "observable indicators about cognition that students create as they engage with a task" (Winne & Perry, 2000, p. 551). Trace file attributes such as time, location, or duration can inform researchers about when and where learners perform specific regulatory activities and how the present SRL activity influences the next SRL activity and how each activity influences learning gains. Compared with other event measures, trace methods are unobtrusive and can make precise inferences about learners' cognitive and metacognitive states in regulation (e.g., Zhou & Winne, 2012)

As a consequence, this paper aims to examine the extent to which self-regulation fosters teachers' deep understandings of TPACK. We will adopt educational data mining and learning analytics techniques to discover the dynamic nature of teachers' regulatory processes. This research addresses the following questions: (1) what SRL subprocesses can be identified with computer trace files while teachers are learning TPACK in CBLEs (2) with the identified SRL processes, what patterns can be discovered about teachers' self-regulation in TPACK learning. The findings from this research can contribute to our scientific understanding of SRL and TPACK learning. It can also make implications for teacher educators and researchers to gain deep insights into teachers' self-regulation, which helps to improve the design of scaffolds in CBLEs to support teachers' learning.

2 RESEARCH METHODOLOGIES

Participants in this research comprise of pre-service (n=50) and in-service (n=20) English teachers from schools and universities in the southern region of China. Pre-service teachers are third-year undergraduate students. Participants consented to participate in the study and were compensated 10 dollars per hour. nBrowser (Poitras, Doleck, Huang, Li, & Lajoie, 2017) - a computer-based

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

learning environment, is used in this research. It is designed to support teachers in learn TPACK by reflecting upon the relationship between technology, content, and pedagogy and internalizing TPACK by asking them to design technology-infused lesson plans. nBrowser consists of two panels (Figure 1). The Workshop Panel has three interfaces. Participants can read the task descriptions, analyze students' characteristics, read about the teaching topics, and refer to the curriculum standards in the Lesson Detail. The Lesson Assets allows participants to search for online information and material through an embedded search engine, saving useful resources using the Bookmark function. In the Lesson Builder, teachers can edit their lesson plans. A virtual agent named Amy was designed in the Tutor session shown at the bottom of the Workspace screen. Amy provided four types of help pertaining to (1) the concept of TPACK, (2) available online resources of educational technologies, (3) a sample technology-infused lesson design, and (4) TESOL criteria of using technology in English teaching. Participants can request Amy's help whenever it is necessary. Once completing the task, participants can review their work and save plans in the Dashboard panel. Each participant has 45 minutes to use the nBrowser to solve problems and complete a lesson plan. nBrowser records participants' interactive actions that take place during TPACK learning and the task solving process, such as opening a web browser, or editing the plan. Besides, participants also complete a survey, reporting their TPACK comprehension before the experiment.

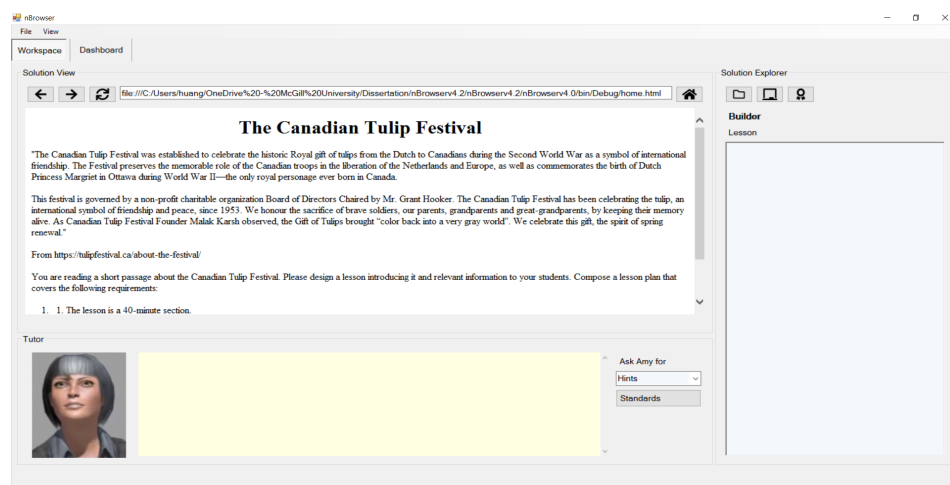


Figure 1 : The screenshot of nBrowser

3 ANALYSIS

Question 1 addresses the identification of teachers' SRL events in nBrowser. We adapted a methodological guide proposed by Siadaty et al. 's (2016) to translate computer log events into indicators of SRL processes. The first step is to establish a reference that identifies SRL processes that could occur in a given learning task in accordance with the SRL model. Table 1 presents the SRL reference applied in this work, which defines the teachers' SRL processes. The left column of Table 1 illustrates four SRL processes on the macro-level. Planning refers to teachers' intentions to get familiar with the teaching tasks and understand the task requirements and goals. Strategy Use refers to the selection of technological tools for instructional purposes and the design of a lesson with the selected tools. Reflection & Adaption emphasizes teachers' self-evaluation and revision processes where they compare their designs against technology use standards and reflect on the effectiveness of their choices. Monitoring pertains to knowing when to request help. The middle column specifies a set of micro SRL activities within each macro phase, which is derived from our previous work

(Poitras et al., 2017). The second step is to develop a pattern library that connects the lower-level events recorded by nBrowser with each of the previous defined SRL events. The elaboration can be found in the right column of the table. For example, the identified *Task Analysis* micro-level process is triggered by two actions: one is that teachers check the Grade drop-down menu to define their students' grade; the other action is checking the Technology competence boxes to report their technological abilities. When the correspondence is established, the third step is to extract features based on trace logs retrieved from nBrowser. The relevant attributes include user ID, Onset time, Elapsed time, Event type, and descriptors. For example, Figure 2 displays the raw computer logs recording information of ID, Time, Events and Event Descriptor. We could find that the participants checked three teaching focus items, one subject topic, and four target language skills that students need to improve, and also defined his/her students' Grade. According to the SRL reference, checking "Topic, Focus, and Language skills" is defined as "goal setting" processes, and checking "Grade" is defined as a "task analysis" process. Consequently, we can extract some features, including two micro SRL processes, i.e., *Task analysis* and *Goal setting*, which also indicates a macro SRL process of *Planning*, as well as the time the participant spends on these SRL processes.

Table 1 Macro- and Micro-level SRL activities and the pattern library

Macro SRL	Micro SRL	Pattern Library
Planning	Task analysis	Clicking on Grade menu, and set students' grade
		Checking technological competencies
	Goal setting	Checking items in the menu of Focus, Topic, and Skills to set goals
		Checking on Standards menu, and set a curriculum goal
Strategy use	Information seeking	Search information in a web browser
	Information acquisition	Assigning tags to a given webpage
		Saving a webpage as a Bookmark
		Adding a new webpage in the Bookmark
Reflection &Adaption	Reflection	Review the Technology use evaluation criteria
		Review the reflective questions
	Adaption	Adding new information to lesson plans
		Removing existing information from lesson plans
Monitoring	Monitor planning	Requesting Hints to understand what technology integration means
		Reviewing the dashboards
	Monitor strategy use	Requesting Hints to open the recommended websites
		Revisiting Home page to coordinate information seeking
		Reviewing the dashboards
		Requesting Hints to open the examples of technology-based lesson designs

userID	Time	MouseLocation	Event	Event Descriptor
20190519001'	'2019.05.19 10:12:19.2539'	'MouseLocation'	'Mouse_Leave'	'Browser_Window'
20190519001'	'2019.05.19 10:12:31.6887'	'Solution_Explorer'	'Lesson_Details_Focus_Checked'	'Theme-based Discussions'
20190519001'	'2019.05.19 10:12:32.8622'	'Solution_Explorer'	'Lesson_Details_Focus_Checked'	'Grammatical Rules'
20190519001'	'2019.05.19 10:12:33.9506'	'Solution_Explorer'	'Lesson_Details_Focus_Checked'	'Vocabulary'
20190519001'	'2019.05.19 10:12:36.2219'	'MouseLocation'	'Mouse_Enter'	'Solution_Explorer'
20190519001'	'2019.05.19 10:12:36.2301'	'MouseLocation'	'Mouse_Leave'	'Solution_Explorer'
20190519001'	'2019.05.19 10:12:36.2773'	'MouseLocation'	'Mouse_Enter'	'Browser_Window'
20190519001'	'2019.05.19 10:12:36.2855'	'MouseLocation'	'Mouse_Leave'	'Browser_Window'
20190519001'	'2019.05.19 10:12:59.9914'	'MouseLocation'	'Mouse_Enter'	'Solution_Explorer'
20190519001'	'2019.05.19 10:12:59.9975'	'MouseLocation'	'Mouse_Leave'	'Solution_Explorer'
20190519001'	'2019.05.19 10:14:10.6869'	'Solution_Explorer'	'Lesson_Details_Topic_Checked'	'Country-related words'
20190519001'	'2019.05.19 10:14:30.4969'	'Solution_Explorer'	'Lesson_Details_Grade'	'First-year university students'
20190519001'	'2019.05.19 10:14:35.2544'	'Solution_Explorer'	'Lesson_Details_LanguageSkills_Checked'	'Listening'
20190519001'	'2019.05.19 10:14:36.3066'	'Solution_Explorer'	'Lesson_Details_LanguageSkills_Checked'	'Speaking'
20190519001'	'2019.05.19 10:14:40.8379'	'Solution_Explorer'	'Lesson_Details_LanguageSkills_Checked'	'Reading'
20190519001'	'2019.05.19 10:14:41.9423'	'Solution_Explorer'	'Lesson_Details_LanguageSkills_Checked'	'Writing'

Figure 2. An excerpt of events extracted from the nBrowser logs

The second research question explores the patterns pertaining to teachers' self-regulation in TPACK learning. For a deep insight, we will classify participants into expert and novice teachers according to their reports of TPACK comprehension and the quality of lesson plans. Process mining (PM) techniques are used to mine the patterns of teachers' regulatory processes for both groups based on the identified SRL events and critical features (Bannert, Reimann, & Sonnenberg, 2014). As fore-mentioned, SRL research has shifted its focus from an aptitude to an event or process perspective (Winne, 2010). Such a perspective posits SRL as a time-stamped but weakly-ordered sequence of regulatory activities. Furthermore, the possibility of recording regulation-related event data in CLBEs satisfies the requirement of PM methods that the data to be analyzed should be recorded event logs (van der Aalst, 2012). Moreover, the PM method has the distinguished advantage of presenting the process in a holistic manner, i.e., how a temporally ordered event sequence is governed by one or more processes (Bannert et al., 2014). It is also time-efficient in mining large event data and providing indices to test the generated model. Therefore, we argue that PM is adequate for investigating regulatory patterns based on process assumptions conceptualized in SRL research. Two PM algorithms were used, fuzzy mining and conformance checking. Fuzzy mining (Günther & van der Aalst, 2007) is used to cluster the unstructured events data into models by providing two interpretable metrics: (a) significance that measures the level of importance of observed events as well as relations and; (b) correlation that measures the relatedness of two events classes (Günther & van der Aalst, 2007). Based on these metrics, the algorithm determines in a process if events are preserved (i.e., highly significant), aggregated (i.e., less significant but highly correlated), or abstracted (i.e., less significant and lowly correlated). Conformance checking (Rozinat & van der Aalst, 2008) is an algorithm that is used to test the validity of the discovered models. The sequence of observed events is tested against a theory-based model represented as a Petri net by examining the extent of fitness (ranges from 0 to 1). In this study, we define the novice model as the observed model and compare it against the expert SRL process model defined as the Petri net to see if there is a difference between them. A higher fitness value suggests a less significant difference between experts and novice while a relatively low fitness can suggest the expert-novice difference and therefore indicate that expert achievements result from success in self-regulation.

4 CURRENT STATUS OF THE WORK

Data collection has been completed in the summer of 2019, and data is being analyzed. Findings from the preliminary analysis are included in pieces of submissions to the annual conference of the American Educational Research Association (AERA) and Learning Analytics and Knowledge (LAK). It is anticipated that the analysis could be finished in early 2020, and one to two related manuscripts could be submitted to a journal.

REFERENCES

- Angeli, C., & Valanides, N. (2009). Epistemological and methodological issues for the conceptualization, development, and assessment of ICT-TPCK: Advances in technological pedagogical content knowledge (TPCK). *Computers & Education*, 52(1), 154–168. <https://doi.org/10.1016/j.compedu.2008.07.006>
- Azevedo, R. (2005). Computer Environments as Metacognitive Tools for Enhancing Learning. *Educational Psychologist*, 40(4), 193–197. https://doi.org/10.1207/s15326985ep4004_1
- Azevedo, R., Guthrie, J. T., & Seibert, D. (2004). The Role of Self-Regulated Learning in Fostering Students' Conceptual Understanding of Complex Systems with Hypermedia. *Journal of*

- Educational Computing Research*, 30(1–2), 87–111. <https://doi.org/10.2190/DVWX-GM1T-6THQ-5WC7>
- Bannert, M., Molenaar, I., Azevedo, R., Järvelä, S., & Gašević, D. (2017). Relevance of Learning Analytics to Measure and Support Students' Learning in Adaptive Educational Technologies. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 568–569. <https://doi.org/10.1145/3027385.3029463>
- Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning*, 9(2), 161–185. <https://doi.org/10.1007/s11409-013-9107-6>
- Greene, J. A., Bolick, C. M., & Robertson, J. (2010). Fostering historical knowledge and thinking skills using hypermedia learning environments: The role of self-regulated learning. *Computers & Education*, 54(1), 230–243. <https://doi.org/10.1016/j.compedu.2009.08.006>
- Günther, C. W., & van der Aalst, W. M. P. (2007). Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics. In G. Alonso (Ed.), *Business process management. 5th International Conference, BPM 2007, Brisbane, Australia, September 24-28, 2007 ; proceedings* (pp. 328–343). https://doi.org/10.1007/978-3-540-75183-0_24
- Kramarski, B., & Michalsky, T. (2010). Preparing preservice teachers for self-regulated learning in the context of technological pedagogical content knowledge. *Learning and Instruction*, 20(5), 434–447. <https://doi.org/10.1016/j.learninstruc.2009.05.003>
- Lajoie, S. P. (2000). *Computers as cognitive tools: No more walls, Vol. II*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Poitras, E., Doleck, T., Huang, L., Li, S., & Lajoie, S. P. (2017). Advancing teacher technology education using open-ended learning environments as research and training platforms. *Australasian Journal of Educational Technology*, 33(3). <https://doi.org/10.14742/ajet.3498>
- Poitras, E. G., Mayne, Z., Huang, L., Udy, L., & Lajoie, S. P. (2018). Scaffolding student teachers' information-seeking behaviours with a network-based tutoring system. *Journal of Computer Assisted Learning*, 35(6), 731–746. <https://doi.org/DOI: 10.1111/jcal.12380>
- Rozinat, A., & van der Aalst, W. M. P. (2008). Conformance checking of processes based on monitoring real behavior. *Information Systems*, 33(1), 64–95. <https://doi.org/10.1016/j.is.2007.07.001>
- Siadat, M., Gasevic, D., & Hatala, M. (2016). Trace-based Micro-analytic Measurement of Self-Regulated Learning Processes. *Journal of Learning Analytics*, 3(1), 183–214. <https://doi.org/10.18608/jla.2016.31.11>
- van der Aalst, W. (2012). Process Mining: Overview and Opportunities. *ACM Trans. Manage. Inf. Syst.*, 3(2), 7:1–7:17. <https://doi.org/10.1145/2229156.2229157>
- Winne, P. H. (2011). A Cognitive and Metacognitive Analysis of Self-Regulated Learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance*. <https://doi.org/10.4324/9780203839010.ch2>
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *The educational psychology series. Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Winne, P. H., & Perry, N. E. (2000). Measuring Self-Regulated Learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). <https://doi.org/10.1016/B978-012109890-2/50045-7>
- Zhou, M., & Winne, P. H. (2012). Modeling academic achievement by self-reported versus traced goal orientation. *Learning and Instruction*, 22(6), 413–419. <https://doi.org/10.1016/j.learninstruc.2012.03.004>

The Importance of Feedback Actionability in Learning Environments

Hamideh Iraj

School of Information Technology and Mathematical Sciences

University of South Australia,

Adelaide, SA, Australia

hamideh.iraj@mymail.unisa.edu.au

ABSTRACT: Feedback is a major factor in student learning in higher education and has been widely researched in the education literature. However, the changes in higher education in the last decades made the provision of feedback challenging. Due to the cut in government funding and the policy to support more students to enter the higher education, large classes were formed in universities in which a large number of students with diverse socio-economic and educational backgrounds were studying. This dramatic shift changed the landscape of feedback; automatic feedback came in as a solution to provide scalable feedback. However, the effectiveness of the feedback process depends on how students perceive the automatic feedback and how different dimensions of automatic feedback affect student learning. Hence the purpose of this doctoral research is to investigate the effect of feedback actionability on academic achievement. It also explores which student sub-populations are more or less likely to act on feedback and how students perceive actionable feedback messages.

Keywords: Feedback, Learning Analytics, Higher Education, Feedback Gap, Data-Driven Approaches

1 INTRODUCTION

Over the past several decades, there has been a continual shift towards massification and consumerisation of higher education. The budget cuts and competition for students have called for efficient and cost-effective operations leading to larger class sizes (Boud & Molloy, 2013; Evans, 2013). However, larger classes are associated with a wide range of unfavourable results such as poor feedback (Gibbs & Simpson, 2005), higher dissatisfaction (Gannaway, Green, & Mertova, 2018), fewer opportunities for monitoring student learning progression (Hattie, 1999) and lower academic performance (Krueger, 1999). Hence, being able to give students quality feedback in a scalable way will help students to have a better learning experience.

Although improving student experience is a complex problem, one of the key factors affecting student learning and success is feedback. This finding is part of a large network of evidence emerging from numerous studies pointing to the power of feedback in education (Hattie & Timperley, 2007). Feedback is only effective if students act upon it (Winstone, Nash, Parker, & Rowntree, 2017). In order to do so, they need to understand the feedback, develop the capacities to judge their work, manage their emotions upon receiving feedback and acting based on the given information (Carless & Boud, 2018). These steps highlight the role of the feedback recipience process (Winstone et al., 2017) in the effectiveness of feedback. A review of findings revealed that what students (not educators) perceive as useful helps students to learn (Price, Handley, Millar, & O'Donovan, 2010; Winstone et al., 2017). However, several of the factors that affect feedback recipience has been currently underexplored in education including feedback actionability, so additional work is required to understand how students perceive feedback and how these factors affect its recipience. The same holds true for technology-mediated feedback; despite the existence of feedback systems such as (Pardo et al., 2018) which provide scalable feedback to students, feedback actionability has not been investigated so far.

This doctoral research aims to investigate the effect of actionability in technology-mediated feedback using learning analytics methods (Siemens & Long, 2011). We would draw on the body of research from digital marketing where message actionability was applied frequently. Specifically, this dissertation will examine the inclusion of "call to action" in student feedback messages and its effect on academic achievement. Cumulatively, the new body of knowledge generated in this dissertation will help the course instructors and designers to develop feedback processes in a way that maximise feedback recipience, and ultimately improve student overall learning experience and success.

2 BACKGROUND

Over the years, two paradigms of feedback have been defined (Carless & Boud, 2018): In the old paradigm, feedback was defined as information. In the seminal paper, Hattie & Timperley (2007, p. 81) defined feedback as "information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding". In the new paradigm of feedback, the focus is on actions rather than information. Feedback goes beyond potentially useful information and is seen as a process to change student behaviour (Carless & Boud, 2018). The focus on actions has been emphasised by (Boud & Molloy, 2013, p. 205) where they defined feedback as "a process whereby learners obtain information about their work to appreciate the similarities and differences between the appropriate standards for any given work, and the qualities of the work itself, to generate improved work".

Feedback by itself is insufficient for improving self-regulated learning and student success (Hattie & Timperley, 2007; Winstone et al., 2017). Understanding feedback can be difficult, so students may need help to comprehend it (Nicol & Macfarlane-Dick, 2006). The discrepancy between the potential and actual uses of feedback has been referred to as a "feedback gap" (Dawson et al., 2018; Evans, 2013) or "feedback paradox" (Withey, 2013). Part of this feedback gap is attributed to the actionability of feedback.

Although the importance of actionability has been acknowledged in the feedback literature (Shute, 2008), the provision of actionable feedback using different educational technologies has been very limited. These include different forms of educational dashboards and email messages, which often did not explicitly include clear actionable information. As shown by (Gašević, Dawson, & Siemens, 2015) not only that the poor design of such systems does not contribute to student success, it can also promote ineffective learning strategies and affect their overall learning negatively. Even when potentially useful information is delivered to students in a technology-mediated manner, the importance of feedback recipience and actionability is often neglected or underestimated (Winstone et al., 2017). For instance, (Corrin & De Barba, 2015) revealed that many students could not interpret their progress shown in educational dashboards, mostly due to confusion and subsequent inaction, making them unable to benefit from the provided feedback. One of the early attempts to collect students' action on dashboards relied on self-reported surveys and interviews (Corrin & de Barba, 2014). Although these research methods might provide useful insights in understanding students' actions, they are not scalable, and they cannot show patterns in bigger cohorts of students.

3 RESEARCH QUESTIONS

The current PhD dissertation addresses the following research questions:

- RQ1: Does the inclusion of actionable link increase student course engagement?
- RQ2: What is the association between student engagement with actionable feedback and their academic success?
- RQ3: What student populations, as captured by demographic and engagement measures are engaging with the provided actionable feedback messages?
- RQ4: How do students perceive actionable feedback – operationalised with CTA trackable links?

The examination of the RQ1 will evaluate whether using actionable links in feedback messages affect student course engagement. By answering RQ2, we can investigate how acting on the provided feedback translates into their learning success i.e. if clicking on actionable links is associated with higher grades. RQ3 investigates which student sub-populations are more or less likely to engage with CTAs and gradually get involved in learning activities and finally, RQ4 investigates how actionable feedback was perceived among students.

4 RESEARCH METHODS

Three courses were selected for this doctoral dissertation: (1) COMP 2016 Digital Media, which is an intensive six-week studio course and a challenging one for many students, (2) NASC 1009 Introduction to biosciences, which is an introductory science course for foundation and diploma students, and (3) MARK 1018 UO Marketing Principles: Trading and Exchange, which is a fully-online course as a part of UniSA online degree program in business.

To answer RQ1, students in COMP 2016 course received personalised email messages in three weeks of the course. Emails in week 4 and 6, included actionable links whereas emails in week 5 did not include actionable links and only provided personalised feedback messages in textual format. By examining the difference in student engagement during these three weeks and the previous offering of the course, we will investigate whether the inclusion of actionable links affects student engagement.

To answer RQs 2-3, we focused on providing students with feedback around their time management in the other two courses. Specifically, students in NASC 1009 received feedback around their engagement with course quizzes while students in MARK 1018 received feedback around their assignments. Each message contained a CTA in the form of a trackable link (See Figure 1) and the content was customised based on students' engagement with previous feedback messages. If students did not engage, they received a reminder a few days later after the original feedback email (See Figure 2).

Hi David,

Well done to review a few key resources for Bioscience! Quiz 2 is coming up soon and it's time to try the practice quiz. This practice will be very helpful for you to get the best marks that you can. Just like before, feel free to attempt the practice quiz multiple times to properly prepare and be ready for Quiz 2.

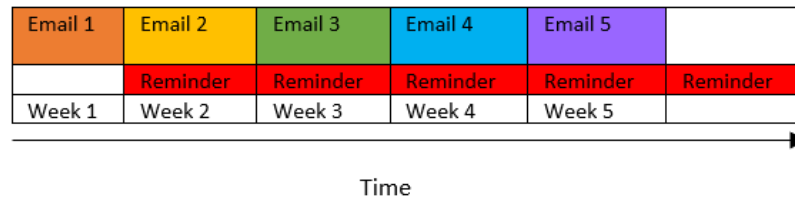
Now that you are aware of the specifics for Quiz 2, you might want to also revisit key resources on the course page that will ensure proper preparation to really do well and maximise your marks.

Any questions or worries please get in contact.

Take Practice Quiz for Quiz 2!

Kind regards,

Figure 1: Example Feedback message

**Figure 2: Structure of feedback provision**

In this thesis, demographics and previous engagement measures were used to predict whether the student engaged with provided feedback using logistic regression (Friedman, Hastie, & Tibshirani, 2001, p. 119). This quantitative method would allow for the iterative improvement of the feedback delivery aligned with design-based research (Design-Based Research Collective, 2003). Moreover, by examining how students engage (and not engage) with the feedback messages, we can refine provided feedback so that it is well-received by the students and used to maximise their engagement in the course. Finally, to answer RQ4 and understand students' perception of actionable links, we ran a few focus groups.

5 CURRENT PROGRESS

The first study for answering RQs 2-4 completed and submitted as a full research paper to LAK 2020. I also presented my research results at HERGA (Higher education research group of Adelaide) 2019. I have two other courses to explore: MARK 1018 – UO Marketing Principles and COMP 2016- Digital Media. Analysing these two datasets will help me in understanding which patterns show up in this course and how similar and different these patterns are compared to the first study. The other manuscript for answering RQ1 is under progress.

Results of the first study revealed that early engagement with the feedback was associated with higher chances of succeeding in the course. Likewise, previous engagement with feedback was highly predictive of students' engagement in the future, and also that certain student sub-populations, (e.g., female students), were more likely to engage than others. Such insight enables instructors to ask “why” questions, improve feedback processes and narrow the feedback gap.

6 DISSERTATION CONTRIBUTIONS

The current dissertation has some major contributions: First, the effect of customised feedback messages has already been investigated qualitatively (Lim et al., 2019). The thesis proposed a new methodology to use trackable links in customised feedback messages to allow quantitative examination of feedback actionability. In other words, in the current thesis, learning analytics methods were used to shed light on the actionability of the feedback messages. Second, after the data has been captured using the data-driven method, the thesis provides empirical evidence for the effectiveness of actionable links in personalised feedback messages across different learning contexts. The patterns found in different learning environments will allow us to see how these patterns resemble or differ among different courses. Third, this dissertation identified opportunities for early identification of students in need of support, so the instructors can design new interventions to help at-risk students to get engaged in the learning process and do not fall behind their peers as the course progresses.

7 SUMMARY

The current doctoral dissertation has several contributions. The researcher can help instructors in designing the interventions, analysing the data and interpreting the results. Instructors will be able to take advantage of the patterns in the data and design subsequent interventions to improve the feedback provision process. A close collaboration between these two groups is essential due to the interdisciplinary nature of the project. Therefore, learning analytics conferences are excellent places to share the results and insights and receive feedback.

REFERENCES

- Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education*, 38(6), 698–712. <https://doi.org/10.1080/02602938.2012.691462>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325. <https://doi.org/10.1080/02602938.2018.1463354>
- Corrin, L., & de Barba, P. (2014). Exploring students' interpretation of feedback delivered through learning analytics dashboards. *Proceedings of the ASCILITE 2014 Conference*, 629–633. Dunedin, NZ.: Rhetoric and Reality: Critical perspectives on educational technology.
- Corrin, L., & De Barba, P. (2015). How do students interpret feedback delivered via dashboards? *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge*, 430–431. <https://doi.org/10.1145/2723576.2723662>
- Dawson, P., Henderson, M., Ryan, T., Mahoney, P., Boud, D., Phillips, M., & Molloy, E. (2018). Technology and Feedback Design. In M. J. Spector, B. B. Lockee, & M. D. Childress (Eds.), *Learning, Design, and Technology* (pp. 1–45). https://doi.org/10.1007/978-3-319-17727-4_124-1
- Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5–8.
- Evans, C. (2013). Making Sense of Assessment Feedback in Higher Education. *Review of Educational Research*, 83(1), 70–120. <https://doi.org/10.3102/0034654312474350>
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). New York, NY, USA: Springer series in statistics New York.
- Gannaway, D., Green, T., & Mertova, P. (2018). So how big is big? Investigating the impact of class size on ratings in student evaluation. *Assessment & Evaluation in Higher Education*, 43(2), 175–184. <https://doi.org/10.1080/02602938.2017.1317327>
- Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71. <https://doi.org/10.1007/s11528-014-0822-x>
- Gibbs, G., & Simpson, C. (2005). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, 1(1), 3–31.
- Hattie, J. (1999). Influences on student learning. *Inaugural Lecture given on August, 2*, 1–25.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Krueger, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics*, 114(2), 497–532. <https://doi.org/10.1162/003355399556052>
- Lim, L.-A., Gentili, S., Pardo, A., Kovanović, V., Whitelock-Wainwright, A., Gašević, D., & Dawson, S. (2019). What changes, and for whom? A study of the impact of learning analytics-based process feedback in a large course. *Learning and Instruction*, 1–11. <https://doi.org/10.1016/j.learninstruc.2019.04.003>
- Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- Pardo, A., Bartimote, K., Shum, S. B., Dawson, S., Gao, J., Gašević, D., ... Vigentini, L. (2018). OnTask: Delivering Data-Informed, Personalized Learning Support Actions. *Journal of Learning Analytics*, 5(3), 235–249. <https://doi.org/10.18608/jla.2018.53.15>
- Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback: All that effort, but what is the effect? *Assessment & Evaluation in Higher Education*, 35(3), 277–289. <https://doi.org/10.1080/02602930903541007>
- Shute, V. J. (2008). Focus on Formative Feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Siemens, G., & Long, P. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*, 46(5), 30–40. <https://doi.org/10.17471/2499-4324/195>
- Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2017). Supporting Learners' Agentic Engagement With Feedback: A Systematic Review and a Taxonomy of Recipience Processes. *Educational Psychologist*, 52(1), 17–37. <https://doi.org/10.1080/00461520.2016.1207538>
- Withey, C. (2013). Feedback engagement: Forcing feed-forward amongst law students. *The Law Teacher*, 47(3), 319–344. <https://doi.org/10.1080/03069400.2013.851336>

Unpacking the Bi-directional relationship between learning analytics and learning design in blended learning environments

Rogers Kaliisa

Department of Education, University of Oslo, Norway
rogers.kaliisa@iped.uio.no

ABSTRACT: This paper suggests a technology-supported teacher-led approach that includes leveraging learning analytics (LA) to support data-informed learning design (LD) decisions in blended learning environments. The context of this study is three to five blended Bachelors courses using the Canvas learning management system (LMS) at two public universities in Norway. This is a design-based research study, employing quantitative ethnography approaches. Data will be collected from multiple sources, i.e. course analytics, discussion forums, interviews, teachers LD representations, and in-class observations. The analysis will be conducted using social and epistemic network analysis, automated discourse analysis, inferential statistics and inductive thematic analysis. This PhD project is anticipated to contribute towards an empirically based theoretical discussion about the potential affordances of LA to transform LD from a craft into a sounder and more evidence-based field of research and practice.

Keywords: Learning analytics, learning design, Canvas, design-based research

1 INTRODUCTION AND BACKGROUND

Learning design (LD), which in this study encompasses “tasks, assessments, learning environments, and resources needed to promote effective interactions between teachers and students and students and students to support learning” (Goodyear & Yang, 2009, p.168), plays an important role in creating an effective learning environment. An LD illustrates the learning objective of a unit of study and is thus useful to teachers and learning designers in supporting them to document their practice (Agostinho, 2011) and improve student learning (Mor, Ferguson, & Wasson, 2015). The common features of all LDs include identifying the key actors (teachers and students), the representations and expectations of each stakeholder (teaching and learning tasks), the resources needed, and the schedule of activities (Lockyer, Heathcote, & Dawson, 2013). Nonetheless, although LD has the potential to highlight pedagogical intentions, it does not always follow an iterative process, which is the hallmark of design and does not take into account how students are engaged in the current course at a fine-grained level of analysis. It also fails to specify the amount of learning that takes place during and after the learning process as specified in the design (Lockyer et al., 2013). Consequently, teachers and learning designers rely on summative assessments (coarse-grained analysis) such as the end of term examinations, course evaluations/surveys, in-class observations, and their previous experience to retrospectively make decisions regarding how best to teach their subjects to the next cohort of students (Persico & Pozzi, 2015). However, with such an approach, little support is given to current students, as changes within the course are only possible and relevant for the next cohort of students (Persico & Pozzi, 2015). Besides, such methods are prone to challenges such as bias, hence providing less objective results (Rienties, Cross, & Zdrahal, 2017). One way to deal with this challenge

is by using more objective and proactive methods to evaluate students' learning in real-time and to enable teachers to make timely informed educational decisions.

Recently, the increasing adoption of education technologies e.g., learning management systems (LMS), online learning approaches such as MOOCs, and content-based learning environments have led to a greater quantity of analyzable learning data and given birth to the field of learning analytics (LA) (Siemens & Long, 2011). These kinds of data, if suitably collected and analyzed, offer more objectivity to the design process by providing immediate feedback and proactive evaluation of students' learning (Persico & Pozzi, 2015). It has been argued that this provides a good base for teachers to make timely, informed educational decisions about redesigning and improving a course and to gain valuable insights into how students react to different learning designs (Nguyen, Rienties, Toetenel, Ferguson, & Whitelock, 2017). The rich and fine-grained data about students' learning behaviours provide teachers with important insights into how students react to different designs, thus allowing educators to make personalized interventions (Rienties et al., 2017). In light of this, within the learning analytics and knowledge community (LAK), there is an increasing interest in exploring the dynamics between LA and LD (Lockyer et al., 2013; Rienties et al., 2017).

2 RELATED RESEARCH AND IDENTIFIED GAPS

The interplay between LA and LD has gained considerable interest over the past few years. For example, Rienties et al. (2017) evaluated the weekly LD data of 2,111 learners in four language studies classes and found that the teachers' course design explained 55% of the variance in weekly online engagement. In another study, Rienties and Toetenel (2016) linked 151 modules taught at the Open University (OU), in which 111,256 students were enrolled, and found that LD was a strong predictor of student satisfaction. A similar approach was taken by Nguyen et al. (2017), who studied 74 modules to examine the impact of assessment design on students' engagement, focusing on fine-grained weekly LD data. Their study indicated that the course workload for other activities diminished after assessment activities were introduced. Moreover, Haya, Daems, Malzahn, Castellanos, and Hoppe (2015) demonstrated the value of an approach that combines social networks and content analysis to support LD decisions by providing indicators that support teachers in their assessment of their LDs.

In another example, McKenney and Mor (2015) argued that the retrospective analysis of LA can support pedagogy-driven data collection and analysis, which could, in turn, offer insight into learning and teaching practices. Meanwhile, Michos, Hernández-Leo, and Albó (2018) more recently explored the connection between LD and data-informed reflection in school environments. Findings from this study indicate that LA was useful in connecting pedagogical intentions and collective reflective practices in school environments.

Recent research has begun to synthesize the corpus of existing research that explores the connection between LA and LD. For instance, Mangaroska and Giannakos (2018) reviewed 43 empirical studies on LA for LD; they depicted ongoing design patterns and detected learning phenomena (i.e. moments of learning or misconception) arising from the connection between LA and LD. Moreover, to aid LA-LD alignment, other research has focused on providing tools and conceptual frameworks to inform the connection between LA and LD (e.g., Bakharia et al., 2016; Hernández-Leo, Martínez-Maldonado, Pardo, Muñoz-Cristóbal, & Rodríguez-Triana, 2019; Lockyer et al., 2013; Persico & Pozzi, 2015;) within online and physical learning settings.

However, and in spite of the increasing interest in exploring the dynamics between LA and LD, my literature review shows that the amount of empirical studies on the subject is still limited. In particular, there is a dearth of evidence to explain how LA is deployed iteratively by instructors to reflect and make informed decisions on their own course designs and to tailor individualized

student support. Also, there is little research on how individual students' fine-grained VLE engagement at the activity level can facilitate the customization of LD. For example, even though research at the Open University (UK) has linked large data sets with students' VLE behaviour, the large log files analyzed make it hard to integrate fine-grained data. Similarly, existing LA and LD studies have not explicitly considered combining digital traces and content-based data (i.e. discussion posts) as a valuable resource in redesigning courses, with inferences only based on trace data (Rienties et al., 2017; Rienties & Toetenel, 2016). Besides, most current LA and LD studies have only been tested in a distance learning setting at one particular, non-traditional university, mainly through applying advanced statistical approaches. An even smaller amount of research (see Michos, Hernández-Leo, & Albó, 2018) has sought information about teachers' experiences with aligning LA and LD based on their generated outputs. The apparent scarcity of studies that use content data and teachers' experiences to acquire a holistic understanding of the connection between LA and LD seems contrary to the documented evidence of utilizing different datasets to offer comprehensive insights and practical comments to support informed future course improvements (Mangaroska & Giannakos, 2018). Therefore, with the motivation to address these research problems, this proposal sketches a study in a traditional blended/face-to-face learning environment using a design-based approach to allow a closer connection of LA with interventionist types of educational research.

3 GOALS AND RESEARCH QUESTIONS

The aim of this doctoral study is threefold. Firstly, to understand the current teacher practice of LD and the perceived potential of LA to support LD decisions across different disciplines at two large public universities in Norway. Secondly, to explore the existing LA models and frameworks, and assess their potential towards LA and LD decisions. Lastly, by building on previous research (Rienties & Toetenel, 2016), my dissertation aims to explore how LA can support data-informed learning design decisions by teachers and how these affect students' learning experiences and performance. The overall research question is: To what extent does a technology-supported teacher-led approach that includes the use of LA help higher education teachers to make data-informed learning design decisions? This question will be investigated through the following specific research questions.

- What is the current teacher practice of LD, and the state of awareness, acceptance, needs and beliefs about applying LA to support learning design decisions?
- What are the features and relevance of the existing LA frameworks in helping teachers to overcome the challenges of LA adoption in their everyday practice?
- What are the opportunities of LA in terms of generating relevant insights about students' online learning processes which teachers can use to make timely and informed pedagogical decisions?
- How can teachers refine, change or adapt the course design while the experience is being delivered using detailed data and representations captured by LA tools and techniques?

4 METHODOLOGY AND ETHICAL CONSIDERATIONS

The central methodological framework guiding this research project is design-based research (DBR). Thus, the study will involve multiple iterations with the aim of understanding possible ways of improving teachers' LD practice through the use of LA. This study will employ quantitative ethnography (QE) approaches such as epistemic network analysis, which will be used to analyze students' online discussions and construct models of student learning that are visualized as network

graphs, and mathematical representations of students' patterns of connections (Shaffer, 2017). The primary sources of data are course analytics data (e.g., activity metrics, and discussion forum posts) collected through the Canvas LMS and representations of teachers' learning designs as visualized on the Canvas LMS. This will be followed by the collection of qualitative data (i.e. interviews with students and teachers) to investigate the implicit meanings/micro-processes and patterns from quantitative analytics data (i.e., why students access certain sites). The student and course weekly statistics will be the unit of analysis to examine student learning behaviour, thus promoting a 'grain size' approach in this study. The study will take place at two large public universities in Norway. The universities offer courses through the traditional face-to-face approach, supported by web-based learning management systems (Canvas) to support face-to-face instruction. To ensure cross-disciplinary representation, the intervention courses will be selected from across the social sciences, arts and science disciplines. Later, results will be compared and aggregated into a body of knowledge to understand the effect of different learning designs on students' learning experiences and performance. It is hoped that this will lead to the identification of good practices in each of the cases and contribute to a community of inquiry (Mor, Ferguson, & Wasson, 2015) at the two universities. While the findings from these two cases may not be generalizable to other contexts, I expect to generate relevant lessons that can be extrapolated with caution elsewhere. To aid the analysis and interpretation of empirical findings, this study will be grounded in a pragmatic, socio-cultural perspective (Knight et al., 2014). Ethical clearance will be obtained by following the Norwegian Centre for Research Data (NSD) and General Data Protection Regulation (GDPR) guidelines.

5 INITIAL FINDINGS

During the first half of my Ph.D., I have conducted four studies in response to the first three research questions. **Study 1:** Current state of LA and LD use: To establish a theoretical basis for my PhD project, I conducted a qualitative investigation with 16 teachers at two large public universities in Norway. The main objective was to understand teachers' current practice of LD, their awareness and perceptions about LA, and whether they perceive the connection between LA and LD useful in their everyday practice. Overall; teachers were positive about LA but also critical about the relevance of the LA outputs and fears of increasing their workload. The findings also revealed that teachers mainly rely on student evaluations, summative feedback and personal experiences to make LD decisions. These findings are in harmony with the core aim of my PhD project which seeks to leverage LA to support teachers with data-informed LD decisions. The main contribution of study one is the proposed Bi-directional LA-LD conceptual framework which considers the synergic relationship between LA and LD (Paper under review).

Study 2: Review of LA frameworks: This paper presents the results of a review of 18 frameworks of relevance to teacher adoption of learning analytics (LA), and discusses how these frameworks have tried to address prominent challenges in the adoption of LA through the lens of relevant literature on the conceptualization of LA adoption. The results show that researchers have made significant advances in developing appropriate frameworks and tools to conceptualize LA adoption at the practitioner level. It was also revealed that LA frameworks have considerably advanced in connecting LA and learning theory. However, the analysis also showed a shortage of explicit guidelines on the required competencies for LA adoption, and strategies to improve inter-

stakeholder communication. Moreover, the review highlights the need to empirically validate, elaborate and put into use the most promising existing frameworks (Paper under review).

Study 3: Exploring social learning analytics (SLA) to inform learning and teaching decisions: This study explored how SLA can be used as a proxy by teachers to understand students' learning processes and to support them in making informed pedagogical decisions. The findings revealed that SLA provides insight and an overview of the students' cognitive and social learning processes in online learning environments. This exploratory study contributes to an improved conceptual understanding of SLA and details some of the methodological implications of an LA approach to enhance teaching and learning decisions in online and blended learning environments (Kaliisa, Mørch, & Kluge, 2019).

Study 4: Combining checklist and process learning analytics to support Learning Design: This study explored the potential of LA to inform LD and how they are experienced by the teachers in a blended learning context. Findings showed that valuable connections between LA and LD require a detailed analysis of students' checkpoint (i.e. online logins see Appendix Fig. 1) and process analytics (i.e., online content and interaction dynamics) to find meaningful learning behaviour patterns that can be presented to the teachers to support design adjustments. Moreover, teachers found LA visualizations valuable to understand students' online learning processes but also argued for the timely sharing of LA visualizations in a simplified interpretable format. The results of this study will be used as input for the next steps (i.e. developing and testing an LA-LD prototype) in authentic learning environments (Paper under review).

6 FUTURE STUDIES

Study 5: Developing an LA prototype for LD with run time application: This phase will involve the development of a research prototype to support the alignment between LA and LD. This will later be applied in real practice to assess the extent to which the detailed data captured by the prototype can support teachers with making real-time adjustments to the LD during the run of the course. The design of the tool will be guided by the insights from studies 1-3, and the recommendations provided by teachers in study 4 (i.e. providing simple LA visualizations to teachers, and hiding unnecessary complexity, but still open to interpretation). This phase will respond to the main PhD research question: How can teachers refine, change or adapt the course design while the experience is being delivered using detailed data and representations captured by LA tools and techniques?

Study 6: Evaluation of the LA-LD prototype: Lastly, an evaluation will be conducted to assess the extent to which the prototype supported teachers with LD decisions. The results from this study will guide future iterations, improvement of the prototype into a learning analytics-learning design tool, and refining of the Bi-directional LA-LD conceptual framework proposed in study one which considers the bi-directional relationship between LD and the LA methods. In other words, the data captured may not only affect LD adaptation but also the type of data to collect and how it is to be structured (i.e. data capturing, sense-making etc.). This means that valuable recommendations for teachers and researchers might be generated.

7 EXPECTED PROJECT CONTRIBUTION

The expected contribution of this PhD project is threefold (i) Conceptually (i.e. developing an empirically grounded LA-LD tool and conceptual framework) (ii) Empirically (contributing towards an

empirically based theoretical discussion about the potential of LA toward informed LD decisions in authentic learning environments, and guidelines to inform practitioners who develop curriculum and technology developers and LA researchers), and (iii) Methodologically (using DBR, quantitative ethnography, and theoretically grounded computational tools). This is an important contribution since rigorous qualitative and design-based research is required to yield actionable insights, provide an explanation for the identified patterns, but also spell out explicitly how LA approaches can be used in different phases of design-based research.

REFERENCES

- Agostinho, S. (2011). The use of a visual learning design representation to support the design process of teaching in higher education. *Australasian Journal of Educational Technology*, 27(6).
- Bakharia, A., Corrin, L., de Barba, P., Kennedy, G., Gašević, D., Mulder, R., & Lockyer, L. (2016, April). A conceptual framework linking learning design with learning analytics. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 329-338). ACM.
- Goodyear, P., & Yang, D. F. (2009). Patterns and pattern languages in educational design. In *Handbook of research on learning design and learning objects: Issues, applications, and technologies* (pp. 167-187). IGI Global.
- Haya, P. A., Daems, O., Malzahn, N., Castellanos, J., & Hoppe, H. U. (2015). Analyzing content and patterns of interaction for improving the learning design of networked learning environments. *British Journal of Educational Technology*, 46(2), 300-316. doi:10.1111/bjet.12264
- Hernández-Leo, D., Martínez-Maldonado, R., Pardo, A., Muñoz-Cristóbal, J. A., & Rodríguez-Triana, M. J. (2019). Analytics for learning design: A layered framework and tools. *British Journal of Educational Technology*, 50(1), 139-152. <https://doi.org/10.1111/bjet.12645>
- Knight, S., Shum, S. B., & Littleton, K. (2014). Epistemology, assessment, pedagogy: where learning meets analytics in the middle space. *Journal of Learning Analytics*, 1(2), 23-47. <https://doi.org/10.18608/jla.2014.12.3>
- Kaliisa, R., Mørch, A. I., & Kluge, A. (2019). *Exploring Social Learning Analytics to Support Teaching and Learning Decisions in Online Learning Environments*. Paper presented at the European Conference on Technology Enhanced Learning. https://doi.org/10.1007/978-3-030-29736-7_14
- Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning Analytics with learning design. *American Behavioural Scientist*, 57(10), 1439-1459.
- Mor, Y., Ferguson, R., & Wasson, B. (2015). Learning design, teacher inquiry into student learning and learning analytics: A call for action. *British Journal of Educational Technology*, 46(2), 221-229.
- Mangaroska, K., & Giannakos, M. N. (2018). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies*. doi: [10.1109/TLT.2018.2868673](https://doi.org/10.1109/TLT.2018.2868673)
- Michos, K., Hernández-Leo, D., & Albó, L. (2018). Teacher-led inquiry in technology-supported school communities. *British Journal of Educational Technology*, 49(6), 1077-1095.
- McKenney, S., & Mor, Y. (2015). Supporting teachers in data-informed educational design. *British Journal of Educational Technology*, 46(2), 265-279. <https://doi.org/10.1111/bjet.12262>.
- Nguyen, Q., Rienties, B., Toetenel, L., Ferguson, R., & Whitelock, D. (2017). Examining the designs of computer-based assessment and its impact on student engagement, satisfaction, and pass rates. *Computers in Human Behavior*, 76, 703-714
- Persico, D., & Pozzi, F. (2015). Informing learning design with learning analytics to improve teacher inquiry. *British Journal of Educational Technology*, 46(2), 230-248.
- Rienties, B., Cross, S., & Zdrahal, Z. (2017). Implementing a Learning Analytics Intervention and Evaluation Framework: what works? In *Big Data and Learning Analytics in Higher Education* (pp. 147-166). Springer International Publishing
- Rienties, B., & Toetenel, L. (2016). The impact of learning design on student behaviour, satisfaction and performance: A cross-institutional comparison across 151 modules. *Computers in Human Behavior*, 60, 333-341.
- Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, 46(5), 30.

Detecting patterns of self-regulated learning in a virtual classroom environment

Madiha Khan

University College London

m.khan.16@ucl.ac.uk

ABSTRACT: There is increasing interest in the use of process mining to track self-regulated learning (SRL) as a dynamic process which unfolds over the course of a learning activity. This form of measurement of SRL is of interest as it could act as the basis for tools which monitor and scaffold SRL in real time in online environments. This doctoral research builds frameworks which theoretically ground the application of process mining in a virtual classroom environment (VCE). It aims to detect the patterns of SRL which exist in practice in this online learning environment, and discuss these in light of existing theoretical models.

Keywords: Self-regulated learning, Process mining, Winne and Hadwin model, Virtual Classroom Environment, analysis framework

1 INTRODUCTION AND CONTEXT

This doctoral research explores the use of Learning Analytics (LA) to detect patterns of self-regulated learning (SRL) in a virtual classroom environment (VCE). A VCE is an online learning environment in which the tutor and student work through online learning resources using natural unstructured dialogue, and an interactive whiteboard. The instruction is conducted on a one to one basis. Tracking patterns of SRL is of interest as it could act as the basis for tools which monitor and scaffold SRL in real time in online environments. The overarching research question is to understand how models of SRL can be applied to inform the use of LA to detect patterns of SRL, amongst high and low performing learners, in a VCE.

Self-regulation has been the subject of extensive research in the past, particularly as it has been strongly linked to academic attainment (Duncan et al., 2007; McClelland et al., 2000). However, a significant portion of this research has been in non-digital environments, with online research becoming more common only in the past couple of decades (Larusson and White 2014). Online research enables self-regulation to be mapped out as a dynamic series of events, which unfold over the course of the learning activity. This is a relatively new perspective on self-regulation, which allows the interaction between the various components of SRL to be examined at a granular level. It contrasts with the traditional approach of measuring SRL as an aptitude i.e. a continuous variable which varies over relatively longer periods of time. Mapping out self-regulation as a dynamic series of events is of interest, as it may provide opportunities for tools which support SRL in real time. Researchers have noted the potential of applying process mining to SRL to identify particular patterns of self-regulated learning (Bannert, Reimann and Sonnenberg, 2014; Behrens and Dicerbo 2014; Pardo 2014; Sedrakyán, De Weerd and Snoeck, 2016; Juhaňák, Zounek and Rohlíková, 2017). However, there is limited research which examines how theoretical frameworks can be used to ground learning analytics. For instance, the most recent systematic review of empirical studies on

Learning Analytics Dashboards (LADs) shows that most of the LADs in the field, %69, were not grounded in any established educational theory [10].

Further, while process mining has been used to detect patterns of SRL using coded think aloud data (Bannert, Reimann and Sonnenberg, 2014) and trace data from online environments in which the learner works independently (Sedrakyan, De Weerdrt and Snoeck, 2016), to the best of my knowledge process mining has not been applied in a VCE. The detected patterns of SRL will build our understanding of how theories of SRL apply in practice in a VCE.

1.1 Structure of this paper

In the remainder of this submission, I will set out the formulated research questions, before providing a brief overview of current research in the problem domains and considering how this has informed the approach to the doctoral work. This will be followed by a brief sketch of the research methodology, before providing an overview of the work conducted to date.

2 RESEARCH QUESTIONS

Title: How can we use models of SRL to inform the use of LA to detect patterns of SRL, amongst high and low performing learners, in a virtual classroom environment?

There are 4 key research questions:

1. How should SRL be conceptualised in a virtual classroom environment (VCE)?
2. What are the indicators of learner self-regulation in a VCE?
3. How can data from a VCE be used to detect patterns of self-regulated learning, amongst high and low performing learners?
4. How can the detected patterns of self-regulated learning be interpreted, in light of existing theories of SRL, and used to advance these theories?

3 STATE OF CURRENT RESEARCH

3.1 Self-regulated learning

There is a diverse body of research on the concept of SRL. While most researchers agree on fundamental aspects of SRL (e.g. striving to achieve goal, actively constructing knowledge, impacted by context), there are also significant theoretical differences in terms of the granularity of the underlying processes and mechanisms e.g. SRL as an event or an aptitude, what the role of context in modelling and scaffolding SRL should be etc (Azevedo *et al.*, 2010).

The doctoral research requires a granular, fluid model of SRL which can be applied to real world settings, and detect fine-grained patterns of SRL from data collected in VCEs. After the review of available theoretical models (Zimmerman, 2000; Greene and Azevedo, 2007; Pintrich 2000a, 2000b; Winne and Hadwin 1998), the Winne and Hadwin [16] model was identified as a suitable model to act as the basis for the research. The Winne and Hadwin model was preferred because i) it has a high level of granularity compared to other models, ii) it is fluid and accommodates learners moving back and forth within phases of self-regulation, and iii) it is heuristic and synthesizes much of the literature on SRL. These features make the theoretical model very suitable for learning analytics research.

The Winne and Hadwin model is depicted in figure 1. The model is structured by component of self-regulated learning rather than by phase, and examines the information processes that occur within each phase of SR through interaction of the SRL components. This structure of the model gives it a higher degree of granularity, and makes it suited to the analysis of fine-grained data that is generated from online environments. Further, the Winne and Hadwin model has noted as being suited to the analysis of online data (Azevedo *et al.*, 2010, Bannert, Reimann and Sonnenberg, 2014) as it synthesises all the various components of SRL from the literature into a heuristic framework that can be applied to educational settings rather than laboratory settings that have been specifically designed to explore SRL. Finally, the Winne and Hadwin model is distinguished by the central role of monitoring in the model. While monitoring and evaluation is seen as a key element of SRL in all models of SRL (Zimmerman, 2000; Greene and Azevedo, 2007; Winne and Hadwin 1998), the Winne and Hadwin model is distinct in that it considers the impact of monitoring and evaluation at a component level, rather than examining impact by phase of SRL (Azevedo *et al.*, 2010). This is a more granular approach, which enables analysis of how changes on one component of SRL (e.g. task conditions) can lead to changes in other components (e.g. type of operations used by learner). The role of monitoring also results in greater fluidity in the model. While all models are weakly sequenced, the Winne and Hadwin model provides a framework for analysing how a learner can move back and forth within a phase of SRL, as well as between phases.

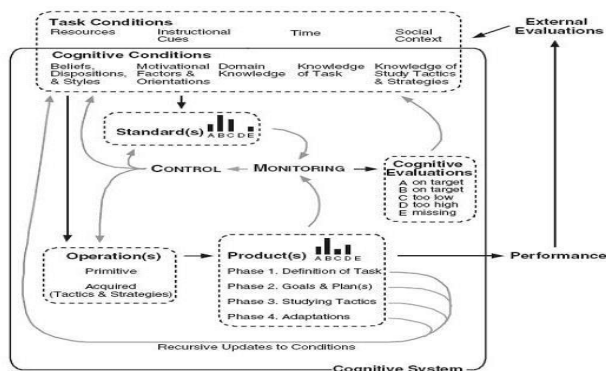


Figure 1: Winne and Hadwin model of SRL

3.2 Process mining for self-regulation

Process Mining aims to discover, track, and refine processes by analysing data from event logs (van der Aalst 2011). Process Mining is applicable when it is assumed that there is a process governing a particular sequence of events (van der Aalst 2011).; this is relevant for learning where learners and tutors can be assumed to engage in particular processes, depending on the specific nature of the task and underlying pedagogy. By reflecting the processual nature of learning, process mining enables process related issues to be highlighted and addressed. For example, bottlenecks within the process which prevent learners from progressing their work has been the focus of previous research (van der Aalst 2011). Furthermore, by discovering and mapping out learner-resource and learner-tutor interactions, it is possible that process mining can also help to identify and understand what types of interactions are related to positive learning outcomes (Juhaňák, Zounek and Rohlíková, 2017). Finally, process mining has the advantage of producing graphical depictions which are

relatively easy for the end-user to interpret and understand (Juhaňák, Zounek and Rohlíková, 2017, Sedrakyan, De Weerdrt and Snoeck, 2016).

In their influential work in the research area, Bannert and colleagues (Bannert, Reimann and Sonnenberg, 2014) led a process mining study on the self-regulation of students working with hypermedia. The researchers coded think aloud verbal data from undergraduate students navigating a hypermedia site, using a theoretical framework specifically designed for hypermedia environments. The researchers selected two groups with high and low attainment scores for further analysis of their self-regulatory processes. A frequency analysis of coded events showed that less successful students relied on shallower learning strategies such as repeating, while more successful students showed a greater frequency of deeper processing, and more diversity in monitoring. Further, the application of fuzzy miner process mining algorithms revealed the maladaptive patterns that less successful students engaged in, such as a loop between monitoring, and reading/repeating. This study demonstrated the value of process mining as a form of learning analytics, in that it allows self-regulated learning to be examined as a process rather than simply examine it as a series of events. Similarly, Sedrakyan, De Weerdrt and Snoeck, 2016 applied an adaptation of Fuzzy Miner called Disco, when analysing the self-regulatory behaviours of computer science students engaging in complex problem-solving task (namely domain modelling). This study demonstrates the flexibility of process mining tools, which can be used to conduct analysis at multiple levels.

The above studies demonstrate the significance and power of process mining when it is based on a theoretical framework tailored for the specific online environment studied. They also highlight the need to merge theoretical frameworks with fine data-driven observations from a broad range of students, as regulatory processes vary significantly with student type.

4 OUTLINE OF RESEARCH METHODOLOGY

The methodology includes 3 key steps (1) development of a framework of indicators (2) validation of the framework through conducting detailed observations and questionnaires for learners, and (3) application of process mining using the framework of indicators to detect patterns of SRL. Stage 1 has been completed, while stage 2 and 3 are in progress.

A mixed methods process is followed to develop the framework of indicators, as per the methodology suggested in Cukurova et al 2016. Firstly, an appropriate theoretical model is identified from the literature. Secondly, the theoretical framework is adapted to fit the research purposes and finally, the fine-grained actions from the data are merged with the adapted theoretical framework. As discussed in 3.1 of this submission, the Winne and Hadwin model has been identified as the preferred model for this research. The model was contextualized to a VCE, by amending the definitions of components of the model to recognize the central role of tutor-student engagement in a VCE. For example, 'monitoring' could include unprompted student monitoring, as well as tutor scaffolding the student to monitor progress. Following the development of the theoretical framework, sessions in a VCE were observed to identify fine grained actions from the data. The VCE used for the doctoral research is being provided by an industrial partner named Third Space Learning (TSL), which delivers one to one maths tutoring, in a virtual classroom environment for year 6 primary school children. Learners and tutors log into a VCE, in which the learner works through a pre-designed online set of questions, with the guidance of his/her human tutor. The learner and tutor are able to write, underline and use a pointer on an interactive whiteboard. Tutors can also leverage motivational tools, including effort points, emojis and pictures. The data available for

analysis includes the online teaching resources, the dialogue between tutor and student, logfile and whiteboard data. No video data is available. 50 sessions from this VCE have been observed on the Number and Place value topic. Fine grained actions that could be observed from the data were recorded and mapped to the theoretical framework of SRL. This exercise illustrated that there were a number of tutor actions aimed at promoting certain types of engagement by learners which were not captured by the theoretical framework. To address this, the 'Operations' component was broadly defined to refer the nature of engagement between the tutor and learner, rather than solely focusing on the student. The final framework developed from this stage of work is attached as an image file.

In the second stage of work, detailed face to face observations will be conducted for 30 sessions. An observation schedule will be developed for SRL events and following completion of the sessions, the online data will be coded using the framework of indicators. The coded data will then be triangulated against the data gathered via observations to validate the framework.

In the final stage of work, the online data for 100 sessions (with identical content) will be coded using the refined framework. The most appropriate process mining algorithm will be selected and applied to the coded dataset to detect patterns of SRL amongst learners and tutors. These patterns will then be explored in light of theoretical models, and the opportunity for real-time tools to promote SRL will be discussed.

To ensure that the empirical work is in line with ethical guidelines, all data that is collected via the online sessions, or the observations and questionnaires will be anonymized and stored securely. Prior to data being collected via observations and questionnaires, the informed consent of schools and learners will be taken.

4.1 Overview of work conducted to date

The first stage of doctoral research referenced in 4 of this submission, namely the development of the framework of indicators has been completed. The final framework is attached.

In addition, a case study has been conducted to test the applicability of the framework using a small sample of online data. The framework was applied to a small sample of 15 online tutoring sessions delivered via VCE. This includes 8 sessions from low attainment learners and 7 sessions from high attainment learners. Learner attainment was defined using a measure of average progress in the content domain. The Average Progress Score (APS) was calculated as the difference between the scores on questions completed at the end of sessions in the content domain, and the scores achieved for that content domain in the diagnostic test after the session. The learner APS was charted, and learners with scores in the 0-10, and 90-100 percentile were obtained. The investigated sessions were selected randomly among the students of these two percentile groups. The selected sessions were coded using the framework, and the results are presented below for Operations and Evaluation events in figure 1 and 2. The bar charts refer to the number of SRL events per category, expressed as a percentage of the total number of SRL events in that category. Although the sample size is too small to draw conclusions from, it is interesting to note that high attaining learners show a higher proportion of diverse monitoring processes and sophisticated operations than low attaining learners. These results are consistent with previous studies on SRL, and point to the applicability and validity of the framework.

In summary, this doctoral research will make two key contributions to the field. Firstly, the work addresses the need for frameworks which can practically act as the basis for LA and tools, as well as

providing theoretical grounding. Secondly, the research will explore the application of process mining to detect patterns of SRL in a VCE. Process mining has been noted as having potential to track SRL in past studies, and this research will build and extend on these findings.

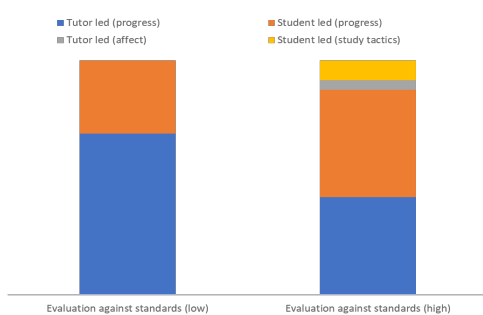


Figure 2: Coded Evaluation events for low vs high attaining learners

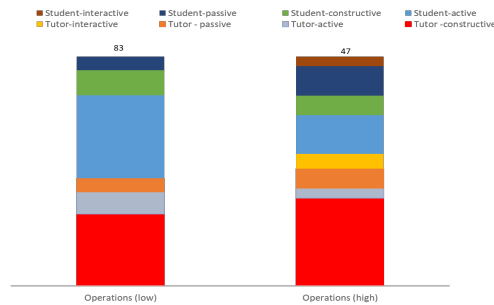


Figure 3: Coded Operations events for low vs high attaining learners

REFERENCES

- Azevedo, R., Moos, D., Johnson, A., & Chauncey, A. (2010) Measuring Cognitive and Metacognitive Regulatory Processes During Hypermedia Learning: Issues and Challenges, *Educational Psychologist*, 45:4, 210-223, DOI: 10.1080/00461520.2010.515934
- Bannert, M., & Reimann, P., & Sonnenberg, C. (2014) Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition Learning* (2014) 9:161–185
- Behrens, J., Dicerbo, K. (2014) Harnessing the Currents of the Digital Ocean. Chapter three, *Learning analytics: from research to practice* / Johann Ari Larusson, Brandon White ;New York, Springer 2014
- Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243
- Cukurova, M, Avramides, K, Spikol, D, Luckin, R, and Mavrikis, M. An Analysis Framework for Collaborative Problem Solving in Practice-based Learning Activities: A Mixed-method Approach (2016): In: *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*. (pp.. 84-88). ACM: New York, NY, USA. (2016)
- Duncan, G.J., Dowsett, C.J., Claessens, A., Magnuson, K., Huston, A.C., Klebanov, P., Pagani, L.S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K. and Japel, C. (2007) School readiness and later achievement. *Developmental Psychology*, 43(6), 1428–46
- Pintrich, P. R. (2000a). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology*, 92, 544–555
- Juhanak, L., Zounek, J., Rohlikova, L. (2019). Using Process Mining to analyse students quiz taking behaviour patterns in a Learning Management System. *Computers in Human Behaviour* (2019) 496-506
- Larusson, J., White, B. (2014). *Learning analytics: from research to practice*. Introduction, p1-12 ; New York, Springer
- Matcha W., Uzir, N., Gasevic, D., Pardo, A. (2019). A systematic review of empirical studies on Learning Analytics Dashboards: A Self-Regulated Learning perspective. *IEEE Transactions on Learning Technologies*. (2019)
- McClelland, M., Morrison, F.J. and Holmes, D.L. (2000) Children at risk for early academic problems: The role of learning-related social skills. *Early Childhood Research Quarterly*, 15, 307–29
- Pardo, A. (2014). Designing Learning Analytics Experiences. In: J. Larusson, B. White eds *Learning analytics: from research to practice*; New York, Springer. Chapter three
- Pintrich, P. R. (2000a). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology*, 92, 544–555
- Sedrakyan, G., De Weerd, J, Snoeck, M. (2016). Process Mining enabled feedback: Tell me what I did wrong vs, tell me how to do it right. *Computers in Human Behaviour* 57 (2016) 352-376
- Van der Aalst, W. (2012). Process mining: Overview and opportunities. *ACM Trans. Manage. Inf. Syst.* 3, 2, Article 7 (July 2012)
- Vandeveld, S., Van Keer, H., Rossell Y.(2013). Measuring the complexity of upper primary school children complexity of Self Regulated Learning; a Multi-Component approach. *Contemporary Educational Psychology* 38 (2013) 407–425
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277-304). Hillsdale, NJ: Lawrence Erl
- Zimmerman, B.J. (2000). Attaining self-regulation, a socio-cognitive perspective. *Handbook of self regulation*. Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

The Institutional Logics of Learning Analytics in Higher Education: Implications for Organizational and Individual Outcomes

Carrie Klein

George Mason University
cklein7@gmu.edu

ABSTRACT: Through a collective case study of a state oversight agency, two large public research universities, and their technology vendors, I conducted 55 interviews to determine the varied institutional logics (e.g., assumptions, values, goals, and outcomes) of higher education learning analytics use in the United States. Initial findings indicate the existence of three dominant logics: technocratic, managerial, and success. These logics have emerged as a result of the organizational demands and social environments in which participant organizations interact. As a sensemaking mechanism, these logics shape specific organizational responses, including changing organizational structures and technological infrastructures, reorienting organizational goals related to student learning and success, and shifting cross-organizational interactions. Further, despite the use of learning analytics to improve student and organizational outcomes, use of these technologies, informed by market, state, and national discourses, have the potential to reify organizational and individual inequities, rather than rectify them.

Keywords: learning analytics, institutional logics, academic capitalism

1 PURPOSE AND BACKGROUND OF THE STUDY

The purpose of my dissertation is to understand the institutional logics of state oversight agencies, technology vendors, and higher education organizations related to learning analytics (LA) use in the United States (U.S.) and how those logics inform subsequent organizational sensemaking and action. Institutional logics are “the socially constructed historical patterns of cultural symbols and material practices, including assumptions, values, beliefs, by which individuals and organizations provide meaning to their daily activity, organize time and space, and reproduce their lives and experiences” (Thornton, Ocasio, & Lounsbury, 2012, p. 2). Institutional logics’ importance lies in their influence over organizational cognition and behavior and in their ability to catalyze sensemaking of and legitimize responses to environmental demands.

I am interested in learning how the increasing presence of LA in higher education influences and is influenced by the institutional logics of state oversight agencies, technology vendors, and higher education organizations. LA are educational big data about learners and their environments. Higher education organizations have adopted LA as a key strategy to improve teaching and learning while at the same time responding to organizational demands – namely accountability and performance metrics (i.e., improved student retention and completion rates) (Long & Siemens, 2011). The proliferation of LA has resulted in a burgeoning educational technology market. For instance, educational technology venture funding, for all levels of education, rose from \$385M in 2009 to \$1.87B in 2014, and investment in educational technology development and deployment by corporate vendors is expected to rise (Koba, 2015). While exact numbers for national spending on LA

in higher education are unclear, the general consensus is that investment in analytics technologies will continue to grow (Ekowo & Palmer, 2016).

2 RESEARCH GOALS AND QUESTIONS

Despite their increasing popularity, the ability of LA to meet organizational demands and improve student outcomes is unclear (Koba, 2015, p. 1; Ferguson & Clow, 2017; Viberg, Hatakka, Bälter, & Mavroudi, 2018). Given the increasing time, money, and resources spent on implementing LA tools with unclear proof of efficacy, it is important to illuminate the institutional logics that govern state oversight agencies, technology vendors, and higher education organizations decision making related to LA. The goals of my research project are to understand the interplay between institutional logics, LA proliferation, and organizational responses and to explain how LA may be shaping not just student learning, but also the structures, interactions, and goals of higher education in the U.S.

The following research questions guided this study: 1) Which institutional logics (i.e., assumptions, values, beliefs, and goals), held by higher education organizations, state higher education governing/oversight boards, and technology vendors, exist related to LA technologies in higher education?; 2) How are these varied logics used to make sense of organizational demands (e.g., the environmental context)?; and 3) How have these logics informed and legitimized subsequent organizational responses and to what end? The focus of these questions is on institutional logics as a sensemaking mechanism for organizational demands and organizational responses. Approaching this study from this perspective allows for greater understanding of where logics, and subsequent sensemaking and response, is aligned or misaligned and legitimized related to LA use.

3 ABBREVIATED LITERATURE REVIEW

A large and ever-growing body of literature has noted the movement of higher education organizations toward market-based structures and practices over the past forty years (Slaughter & Rhoades, 2004). Since the 1980s, academic institutions in the U.S. have adopted more corporatized forms and initiatives as a result of increased oversight by state regulatory agencies and increased interaction with market actors, what Slaughter & Rhoades (2004) refer to as *academic capitalism*. Academic capitalism has resulted in increased accountability measures by state oversight agencies, increased vendor involvement in traditional academic spheres, and the proliferation of market-based priorities and behaviors of higher education organizations (Olssen & Peters, 2005; Slaughter & Rhoades, 2004).

In response to state and market demands, technological innovations have become a key catalyst of change in higher education (Slaughter & Rhoades, 2004). LA are an example of such an innovation. The value of LA lie in their potential to improve organizational capacity for data-informed decision making through the mining of vast amounts of varied data that provide timely and predictive feedback for both students and higher education organizations (Norris & Baer, 2013). LA are viewed as one way for organizations to respond, “to internal and external pressures for accountability in higher education, especially in the areas of improved learning outcomes and student outcomes” (Norris & Baer, 2013, p. 11) from a conceivably more informed, dynamic, and efficient perspective. As a result, LA have been deployed to improve course and curriculum design and assessment, student learning, engagement, and performance, and evaluation of policies and academic quality (Dahlstrom, Brooks & Bichsel, 2014; MacFadyen & Dawson, 2012).

However, evidence of LA's efficacy to improve student learning and performance is murky, and critics argue that these tools are "not yet a proven solution for learning and limit the experience of education and human interaction" (Koba, 2015, p. 1). Further, the popularity and potential of LA does not come without limitations and concerns, including a lack of user interest, knowledge, and time (Macfadyen & Dawson, 2012), institutional readiness and capacity (Norris & Baer 2013; Oster, Lonn, Pistilli, & Brown, 2016), technological alignment to users' needs (Klein, Lester, Rangwala, & Johri, 2019), and limited empirical evidence that LA, when deployed at scale, can transform student learning and organizational functioning (Viberg, Hatakka, Bälter, & Mavroudi, 2018). Importantly, concerns tied to social justice, privacy, and ethics also abound, including: algorithmic bias and discrimination (Diakopoulos, 2015; Hajian, Bonchi, & Castillo, 2016); data trust and transparency (Pardo & Siemens, 2014; Slade & Prinsloo, 2013; Rubel & Jones, 2016); data governance, security and access (Drachsler & Greller, 2016; Johnson, 2018; Steiner, Kickmeyer-Rust & Albert, 2016); and data rights (Beattie, Woodley, & Souter, 2014; Johnson, 2018). Despite these challenges, higher education is increasingly moving toward data-informed decision making, driven by emerging LA use.

4 NOVELTY OF RESEARCH APPROACH

Numerous studies have been conducted to understand LA tool development, including a focus on efficacy of data visualizations and interventions on user behaviors and on organizational use of these tools (Viberg, Hatakka, Bälter, & Mavroudi, 2018). However, empirical research focusing on the social implications of this work is limited. Recent literature has focused on the ethics, privacy, and policy concerns related to LA tools (Rubel & Jones, 2016; Drachsler & Greller, 2016; Pardo & Siemens, 2014; Prinsloo & Slade, 2015, 2013; Selwyn, 2016, 2015; Slade & Prinsloo, 2013; Willis, Slade & Prinsloo, 2016) and on the perspectives, needs, and interests of those using these tools (Aguilar, 2018; Klein, Lester, Rangwala & Johri, 2019a, b). Though work has been done to explore the varied institutional logics of LA technologies between various institutional involved in the proliferation of these tools in Australia (Dawson, et al., 2016), this study is among the first of its kind in the U.S. to explore these technologies and interactions from this perspective.

5 METHODOLOGY

To better understand the institutional logics of LA use in higher education in the United States, I conducted a collective case study focused on the interactions between state oversight agencies, technology vendors, and higher education organizations within a single state university system. Collective case studies allow for the investigation, description, and comparison of multiple, related cases, "that provide insight into an issue, theme, or phenomenon (Creswell, 2012, p. 477). The phenomenon of focus is the institutional logics of these organizations related to LA use. The cases involved in this study exist in the southeastern U.S. Participant organizations included the academic affairs unit of the state oversight agency, two public research universities (one rural flagship institution and one urban access institution), and one technology vendor associated with each university. The state oversight agency was encouraging use of analytics data¹ across its centralized

¹ Participants from state oversight agencies and higher education organizations did not differentiate between learning analytics data and other type of analytics data (e.g., business analytics, academic analytics, or advising/success analytics), viewing data as data, regardless of where it was harvested or how it was recombined, and using that data to improve organizational and student outcomes. However, vendors differentiated analytics data (ostensibly as a marketing function), focusing on success analytics, regardless of which data were feeding their algorithms.

system to encourage improved student and university outcomes. The universities had partnered with vendors and were actively engaged in using analytics data to inform institutional efforts. After obtaining ethics board approval in Spring 2019, I conducted 55 semi-structured interviews with members of the participant organizations and reviewed associated artifacts (e.g. websites, reports, white papers, etc.). In Summer 2019, I completed interview transcription and began data coding and analysis. I open coded the transcripts, creating connections and themes, from which I derived the three dominant institutional logics. Trustworthiness and validity are being addressed using member checking and triangulation.

6 CURRENT STATUS

I defended my proposal in Spring 2019 and finished collecting, coding, and analyzing data in Summer 2019 and am currently writing my findings. Initial findings include the existence of three dominant institutional logics: technocratic, managerial, and success. The technocratic logic is tied to the potential of LA, rather than current capabilities. Participants value LA's potential to inform decision making, but their technological infrastructures, data quality, and data literacy often limit the efficacy of LA. Regardless, the technocratic logic has legitimized an increased focus on data collection to support organizational decision making. The strength of the managerial logic at play is also driving LA use through the restructuring and reorienting of organizational structures, resources, personnel, and programming. These organizational changes are closely tied to the types of data and technologies available to organizations, meaning that organizational processes are changing and being created to support specific LA technologies available to an organization. Moreover, the managerial logic also leverages market-associated language to communicate the need for both LA and organizational change to support LA use with the ultimate goal of improving student and organizational outcomes. Among the strongest discourses at play related to outcomes is a focus on success – for both students and organizations. The success logic permeated all organizations, is tied to organizational mission, identity, and context, and is used to leverage LA to better meet student needs and organizational priorities. However, variable organizational contexts, specifically between the participant universities, can also create unintended organizational and individual inequities. For example, LA are used at the access institution for course, credentialing, financial aid support, and pathway completion with an eye on workforce development. Conversely, at the flagship institution, while they track similar outcome metrics, LA are used to, as a participant described, “optimize the student experience” by maximizing their experiences (e.g., steering students to study abroad, internships, mental health support, etc.). Although, these organizations are meeting the needs of their students within their specific contexts, the outcomes for these students are still decidedly different. Students at the access institution may well get a degree and improve their individual outcomes, but students at the flagship are still graduating at higher rates, with better economic outcomes with an additional ‘optimized’ experience. Further, the success logic legitimizes increased surveillance of students’ interactions with their institutions. These findings indicate that the institutional logics of LA play a role in sensemaking, decision making, and outcomes, with important implications for higher education organizations and their students. Among the unique contributions of this study is its focus on U.S. higher education organizations across different fields and an interrogation of assumptions about improved equity and outcomes via LA-supported success initiatives. The work closes with recommendations for ways to disrupt logic-based assumptions associated with LA use. As I continue to write my findings and my discussion and implications chapters, I would welcome input from doctoral consortium members and LAK2020/SoLAR scholars.

REFERENCES

- Aguilar, S. J. (2018). Learning analytics: At the nexus of big data, digital innovation, and social justice in education. *TechTrends*, 62(1), 37-45.
- Beattie, S., Woodley, C., & Souter, K. (2014). Creepy analytics and learner data rights. *Rhetoric and Reality: Critical Perspectives on Educational Technology. Proceedings ascilite*, 421-425.
- Creswell, J.W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*, 4th edition. New York, NY: Pearson Publishing.
- Dahlstrom, E., Brooks, D. C., & Bichsel, J. (2014). *The Current Ecosystem of Learning Management Systems in Higher Education: Student, Faculty, and IT Perspectives* (Research report). Louisville, CO: EDUCAUSE, September 2014.
- Dawson, S., Wade, A., Rodgers, T., Gasevic, D., Buckingham Shum, S., Alexander, S., Lockyer, L., Nelson, K., Kennedy, G., Corrin, L., & Fischer, J. (2016) Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement, final report prepared for the Australian Government, Office for Learning and Teaching.
- Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, 3(3), 398-415.
- Drachsler, H., & Greller, W. (2016, April). Privacy and analytics: it's a DELICATE issue a checklist for trusted learning analytics. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 89-98. New York, NY: Association for Computing Machinery (ACM).
- Ekowo, M., & Palmer, I. (2016). *The promise and peril of predictive analytics in higher education*. Washington, DC: New America.
- Ferguson, R., & Clow, D. (2017, March). Where is the evidence?: A call to action for learning analytics. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 56-65). Vancouver, Canada. ACM.
- Hajian, S., Bonchi, F., & Castillo, C. (2016, August). Algorithmic bias: From discrimination discovery to fairness-aware data mining. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2125-2126. New York, NY: Association for Computing Machinery (ACM).
- Johnson, J. A. (2018). *Toward information justice: Technology, politics, and policy for data in higher education administration* (Vol. 33). New York, NY: Springer.
- Klein, C., Lester, J., Rangwala, H., & Johri, A. (2019a). Learning analytics tools in higher education: Adoption at the intersection of institutional commitment and individual action. *The Review of Higher Education*.
- Klein, C., Lester, J., Rangwala, H., & Johri, A. (2019b). Student sensemaking of learning analytics dashboard interventions. *Journal of Educational Technology Systems*, 48(1), 130-154.
- Koba, M. (2015, April 28). Educational technology soars—but is it working in the classroom? Fortune Magazine. Retrieved from: <http://fortune.com/2015/04/28/education-tech-funding-soars-but-is-it-working-in-the-classroom/>.
- Long, P. & Siemens, G. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*, 46(5), 30.
- Macfadyen, L. P., & Dawson, S. (2012). Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. *Journal of Educational Technology & Society*, 15(3), 149-163.

- Norris, D. M., & Baer, L. L. (2013). *Building organizational capacity for analytics*. Louisville, CO: EDUCAUSE. Retrieved from <http://www.educause.edu/library/resources/building-organizationalcapacity-analytics>.
- Olsen, M., & Peters, M. A. (2005). Neoliberalism, higher education and the knowledge economy: From the free market to knowledge capitalism. *Journal of Education Policy*, 20(3), 313-345.
- Oster, M., Lonn, S., Pistilli, M. D., & Brown, M. G. (2016, April). The learning analytics readiness instrument. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 173-182). ACM.
- Papamitsiou, Z. K., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society*, 17(4), 49-64.
- Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438-450.
- Peters, M. A. (2006). Higher education, development and the learning economy. *Policy Futures in Education*, 4(3), 279-291.
- Prinsloo, P., & Slade, S. (2013, April). An evaluation of policy frameworks for addressing ethical considerations in learning analytics. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 240-244. New York, NY: Association for Computing Machinery (ACM).
- Prinsloo, P., & Slade, S. (2015, March). Student privacy self-management: implications for learning analytics. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge*, 83-92. New York, NY: Association for Computing Machinery (ACM).
- Rubel, A., & Jones, K. M. (2016). Student privacy in learning analytics: An information ethics perspective. *The Information Society*, 32(2), 143-159.
- Selwyn, N. (2015). Data entry: towards the critical study of digital data and education. *Learning, Media and Technology*, 40(1), 64-82.
- Selwyn, N. (2016). *Is technology good for education?* Hoboken, NJ: John Wiley & Sons.
- Slade, S., & Prinsloo, P. (2013). Learning analytics ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510-1529.
- Slade, S., & Prinsloo, P. (2015). Student perspectives on the use of their data: between intrusion, surveillance and care. *European Journal of Open, Distance and E-learning*, 18(1), 291-300.
- Slaughter, S., & Rhoades, G. (2004). *Academic capitalism and the new economy*. Baltimore, MD: Johns Hopkins University Press.
- Steiner, C. M., Kickmeier-Rust, M. D., & Albert, D. (2016). LEA in private: A privacy and data protection framework for a learning analytics toolbox. *Journal of Learning Analytics*, 3(1), 66-90.
- Thornton, P. H., Ocasio, W., & Lounsbury, M. (2012). *The institutional logics perspective: A new approach to culture, structure, and process*. Oxford, England: Oxford University Press.
- Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, 57(10), 1500-1509.
- Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior*, 89, 98-110.
- Willis, J. E., Slade, S., & Prinsloo, P. (2016). Ethical oversight of student data in learning analytics: a typology derived from a cross-continental, cross-institutional perspective.

Understanding Metacognition in Online Learning: To What Extent Learners Trajectories and Self-Reported Behaviors Converge

Fatemeh Salehian Kia¹, Marek Hatala¹, Ryan S. Baker²

¹ Simon Fraser University, ² University of Pennsylvania
fsalehia, mhatala@sfu.ca, ryanshaunbaker@gmail.com

ABSTRACT: Research has identified self-regulated learning as an effective approach to learning. However, much less understood about how to detect self-regulated learning behaviors in online learning in order to support students. This proposal presents a method to detect and test the measures of metacognitive behaviors in the open-ended learning environment (i.e. LMS). The log tracing technique allows to track students' online activities. We will develop proximal indicators of SRL behaviors from students' logs of interactions in the LMS. However, the affordances and limitations of LMS restrict capturing all aspects of students' metacognition. In addition, students can be inaccurate in calibrating their self-reported metacognitive behaviors. This doctoral dissertation proposes to triangulate two sources of ground truth for behavior detection i.e. students' trace of learning behaviors and self-report of metacognition to mitigate the limitations of behavioral data and test the validity of inferences about students' self-regulatory behaviors.

Keywords: Self-regulated learning, proximal behavior indicators, self-reported behaviors

1 INTRODUCTION

Self-regulated learning is an important component of learning for students as research emphasizes that it is critically related to academic performance (Schunk & Green, 2017). However, students differ in their self-regulatory skills; they have different capabilities, know and adopt different strategies and tactics to study. Research showed that students can learn how to become self-regulated learners, and educators can foster self-regulatory skills in their courses (Pintrich, 1995). Supporting students, who lack self-regulatory skill or engage in undesirable learning behaviors, requires identifying students' current learning behaviors. Increasing use of technology in education and the resulting ability to keep track of learning activities offer the opportunity to understand students' behaviors at an unprecedented level of analysis and scale. A few studies have adopted methods such as sequence analysis or process mining (Gasevic et al., 2017; Jovanović, 2017; Bannert et al., 2014) to detect SRL behaviors from students' trace data in online learning environments.

Despite recent advances in detecting students' metacognitive behaviors, there has been much less research on understanding to what extent the trace of learners' online activities (known as proximal behavior indicators) contribute to valid inference of their self-regulated learning behaviors. In addition, the methods employed to identify behavior indicators need to be tested for the extent and conditions to infer SRL behaviors. A recent study by Matcha et al. (2019) compared the results of three data analysis approaches (i.e. sequence, process, and network analytics) to detecting learning tactics and strategies in a MOOC. Their findings showed differences in detecting learning tactics from the same trace data between three different methods. In this doctoral dissertation, I propose to design a learning experience in our learning management system (Canvas) that makes it feasible to

observe students' activities in order to develop and test proximal indicators that contribute to valid inferences of self-regulatory behaviors.

2 MOTIVATION

Research on SRL measures has shown that learners can be inaccurate in calibrating their self-reported SRL behaviors (Rovers et al., 2019; Zhou et al., 2012). Since disagreements exist regarding SRL measurement, particularly whether self-reports represent a valid and reliable approach to measure this process, researchers have advocated the use of behavioural measures of SRL (Zhou et al., 2012). Therefore, detecting latent constructs of learning behavior such as metacognition from learners' trajectories of online activities provides promising insight on the learning process besides traditional approach to measuring self-regulated learning by self-report questionnaires. A study by Winne and Jamieson-Noel (2002) is among early research on collecting students' trace data to measure self-regulatory skills. They used log data of software called *PrepMate* designed for authoring instructional presentations to capture students' interactions with reading materials. For instance, scrolling through a paragraph in the reading material before starting any other actions was used as a proximal indicator of students' SRL planning skill, whereas a recent study by Cicchinelli et al. (2018) used students' access to resources related to course organization such as exercise deadlines or index page with learning objectives in the LMS as proximal indicators of students' SRL planning skill. Although these studies conducted in two different contexts to detect students' SRL behaviors, these proximal indicators of SRL planning skill are quite variable in terms of operationalization. The former used a learning tactic (i.e. skimming) as students' planning skill while the latter used indicators identifying students' more global skill set of planning (i.e. organization and goal setting). Thus, there is a need of more consistent approach to developing proximal indicators of SRL behaviors and operationalizing students' behaviors in online learning.

Furthermore, capturing features of learners' behavior context is more feasible when learning activities occur in well-defined tasks in online or computer-based environments rather than open-ended learning environments such as LMS or MOOCs. Most prior research on studying SRL behaviors was conducted within well-defined tasks in tools such as *gstudy*, a cross-platform software tool for researching learning. Hadwin and colleagues (2007) conducted an exploratory case study to examine in depth the students' learning activities across studying episodes in *gstudy*. But examining a studying episode in open-ended environments, particularly LMS, is not the same because of content-agnostic nature of LMS that makes it more difficult to capture the context of students' actions, in addition to a lack of action end time logging. In fact, similar observable activities in the LMS may represent different behavior in different learning context, which is an important component of self-regulated learning (Roll et al., 2015). Recent research on detecting SRL behaviors in open-ended environments has been shifted from using frequency of students' activities to observing students' sequences of actions in order to better understand the context of students' choice of actions. A study by Jovanović et al. (2017) is an example of adopting the combination of exploratory sequence analysis and hierarchical clustering to detect patterns in students' behaviors that are indicative of the adopted learning strategies in the LMS. However, without considering course sitemap in the LMS, some sequences of actions may be misinterpreted as indicator of the same behavior and grouped in the same cluster. For instance, tabbed browsing allows students to click on multiple links within a few seconds and open multiple pages by loading them into tabbed sections of one page.

These multiple (almost simultaneous) access actions may interpret in the same order that they are opened while we cannot inevitably conclude either this sequence of actions in such order or all those pages viewed.

Tracing metacognitive processes, which are not directly observed from learners' trajectories in online learning, is an inferential pursuit (Bernacki et al., 2015). In addition, due to the specific affordances of LMS, we only can track a glimpse of what occurs outside the LMS. The trace data as a source of ground truth for metacognitive behavior is thus limited. We must wrestle with more noise and indeterminacy when we detect metacognitive behaviors from logs in the LMS than in well-defined learning systems. Therefore, another source of ground truth can help to make more valid inference of behaviors. Among sources of ground truth for behavior, a self-report questionnaire is a feasible approach to collect data in the LMS and test the validity of proximal behavior indicators. Moreover, there is evidence that converging self-report and process data is key to understanding metacognitive self-regulatory processes (Mudrick et al., 2019).

Given the problems explained above, this doctoral dissertation contributes to the growing research of SRL modeling in several ways. First, we intend to **design a learning activity**, asking how a task can be designed that makes it possible to observe students' interactions with information provided to perform a task in the LMS. Second, we will **develop proximal indicators** of self-regulated learning process from students' trace data. Third, we will **implement a prompting system** to collect students' self-reported SRL and minimize the concerns associated with students reporting their own SRL behaviors. Finally, based on both log and self-report data, our research design allows us to examine the extent to which proximal indicators **converge** with their corresponding self-reported SRL measures to mitigate our data limitations.

3 RESEARCH QUESTIONS

The present dissertation will be guided by the following research questions:

***RQ1.** How can we detect proximal indicators of SRL behaviors in real-time using log tracing techniques, given the specific affordances and limitations of the LMS?*

***RQ2.** What specific schedule leads to collect the best students' self-reported SRL behaviors to minimize the concerns associated with self-reporting metacognitive behaviors?*

***RQ3.** To what extent do proximal indicators detected from log data converge with self-reported behaviors?*

4 METHOD

4.1 Designing a learning activity

We intend to design a learning activity in the LMS that allows us to track students' access to information provided for them to perform a task. We chose assignment module because it is the most commonly used module among students and instructors in the LMS, which can generate richer log data of students' interactions. In addition, the assignments are mandatory and required to submit within a time frame of two weeks after they are published online in the LMS. The final course

assessment is influenced by the students' performance on assignment tasks. The tasks aim to promote understanding about basic concepts of object-oriented and event-driven programming by practicing these concepts through hands-on experiences. The students are provided with instructions and additional information about how to perform a task and hand it in through the assignment module in the LMS. Since the students work throughout a programming task offline and upload their solutions as assignment submissions in the LMS, they are asked to complete a follow-up quiz about programming concepts targeted in the task to test students' programming proficiency.

The proximal indicators of metacognitive behaviors will be developed based on students' access behavior to different kinds of information offered as task instructions. The instructions consist of six different kinds of information (i.e. *Overview*, *General Guideline*, *Detailed Specification*, *Resources*, *Marking Criteria*, and *How to Submit*). Each part of instructions aims to support students' learning process to perform a task, and we are also able to track what information they access. The *Overview* page provides the learning objectives of a task. The *General Guideline* section offers steps that can be followed to perform a task. The *Detailed specification* describes step-by-step instructions on task requirements. In *Resources* page, students are provided with the list of supplementary resources e.g. links to lecture notes and worked examples related to the task. The *Marking Criteria* page clarifies how the task will be evaluated and *How to Submit* page instructed students on submission formats and deadlines. A graphic organizer—inspired by the concept of process flowcharts-- includes links to all instruction pages in assignments module (An example is shown in Figure 1).

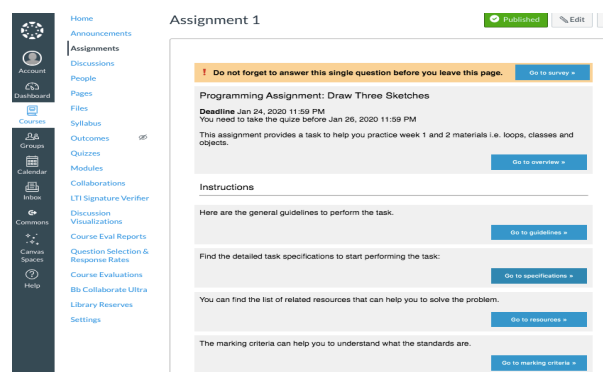


Figure 1: The graphic organizer representing assignment task instructions

4.3 Developing proximal indicators of SRL behaviors

Interaction data was collected from 92 undergraduate students who enrolled in a computer-programming course offered in the Spring 2019 term, at a residential comprehensive research university in Canada. The assignment was available to these students for two weeks of the term. All students' fine-grained actions were time-stamped and recorded by the LMS. These actions included: interactions with the course pages, interactions with the graphic organizer in the assignments module including viewing task instruction pages (i.e. *Overview*, *General Guideline*, *Detailed Specification*, *Resources*, *Marking Criteria*, and *How to Submit*), accessing the learning resources, and online discussion participation. We are in the process of distilling raw logs and find indicative sequences of these actions representing self-regulatory learning behaviors. Contiguous sequences of these actions are segmented into *clips*. A *clip* contains all the actions in the order necessary for a

human coder to identify whether students demonstrate self-regulated learning behaviors. *Clips* are the grain-size at which self-regulatory behavior is detected. We continue building *clips* until there is no new one found in raw logs. From there, indicators describing what it means to demonstrate SRL behaviors in an assignment are developed.

For our domain, the lens we use to detect SRL behaviors is the well-established model of SRL by Winne and Hadwin (1998) from information processing perspective since we detect students' behaviors from interactions with information offered as task instructions to them. This model of SRL defined loosely sequence cycle of four phases to perform an academic task: (1) task definition, (2) planning and goal setting, (3) enactment of tactics and strategies, (4) adapting. In each of these phases, learners find themselves in a set of processes involving interaction between the conditions, operations, products, evaluations and standards (COPEs) (Winne, 2017). Next, we determine whether or not students demonstrate one of four SRL phases within clips, by labeling *text replays of clips* with one or more tags (see Sao Pedro et al. (2013) for an example of a text replay). The clips having all actions taken while formulating a task including access to information about task **conditions** and learning objectives are tagged as "*task definition*", setting goals including first time access to information concerning task **operations** are tagged as "*planning*", writing a computer program in IDE including access to information regarding task **operations** are tagged as "*enactment*", and making **evaluations** of the computer program and adapting programming **product** to the task requirements including access to information about assessment criteria or task **standards** are tagged as "*adapting*".

4.4 Collecting self-reported SRL data

Self-report questionnaires are a form of SRL measures that is easy to collect in the LMS, however, there are two main concerns regarding self-reported SRL measures. First, SRL is considered to be a context-dependent process (Zhou et al., 2012), and it may vary both across and within learning tasks and contexts. The second concern is the question of whether students have the capacity to self-report their own metacognitive behaviors because of students' imperfect memory, confusion of several contexts, and students' inclination to provide socially desirable answers (Rovers et al., 2019). Due to these concerns, researchers have increasingly advocated the use of other forms of SRL measures, that is, a multi-method approach such as incorporating short micro-analytic questionnaires at various points during a learning episode (Cleary et al. 2015).

To address the concerns associated with self-reported SRL measures, we will conduct a quasi-experiment in the same undergraduate computer-programming course in the Spring 2020 term. Participation in the study will be voluntary. The participants will be asked to report their self-regulatory behaviors at various points during performing the assignment. We will implement a prompting system asking the participant a single-item question to report on their behavior when a behavior indicator is observed in the participant's logs. We will determine the best schedules of prompting procedure based on our findings in the previous study.

4.5 Converging proximal indicators and self-reported data

To address *RQ3*, we will examine whether there is an association between the learner's trajectories of self-regulated learning, represented by indicative sequence of actions (described in Subsection

4.3), and their self-reported SRL behaviors. The analysis will be carried out on indicative sequences for total number of participants. A set of sign tests will be conducted to identify indicative sequences (proximal behavior indicators) that are associated with the corresponding self-reported SRL constructs (i.e. task definition, planning, enactment, and adapting).

REFERENCES

- Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and learning*, 9(2), 161-185.
- Bernacki, M. L., Nokes-Malach, T. J., & Aleven, V. (2015). Examining self-efficacy during learning: variability and relations to behavior, performance, and learning. *Metacognition and Learning*, 10(1), 99-117.
- Gasevic, D., Jovanovic, J., Pardo, A., & Dawson, S. (2017). Detecting learning strategies with analytics: Links with self-reported measures and academic performance. *Journal of Learning Analytics*, 4(2), 113-128.
- Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, 2(2-3), 107-124.
- Jovanović, J., Gašević, D., Dawson, S., Pardo, A., & Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *The Internet and Higher Education*, 33(4), 74-85.
- Matcha, W., Gašević, D., Jovanović, J., Pardo, A., Maldonado-Mahauad, J., & Pérez-Sanagustín, M. (2019, September). Detection of Learning Strategies: A Comparison of Process, Sequence and Network Analytic Approaches. In *European Conference on Technology Enhanced Learning* (pp. 525-540). Springer, Cham.
- Pintrich, P. R. (1995). Understanding self - regulated learning. *New directions for teaching and learning*, 1995(63), 3-12.
- Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, 2(1), 7-12.
- Rovers, S. F., Clarebout, G., Savelberg, H. H., de Bruin, A. B., & van Merriënboer, J. J. (2019). Granularity matters: comparing different ways of measuring self-regulated learning. *Metacognition and Learning*, 14(1), 1-19.
- Sao Pedro, M. A., Baker, R. S., & Gobert, J. D. (2013, April). What different kinds of stratification can reveal about the generalizability of data-mined skill assessment models. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 190-194). ACM.
- Schunk, D. H., & Greene, J. A. (2017). *Handbook of self-regulation of learning and performance*. Routledge.
- Mudrick, N. V., Azevedo, R., & Taub, M. (2019). Integrating metacognitive judgments and eye movements using sequential pattern mining to understand processes underlying multimedia learning. *Computers in Human Behavior*, 96, 223-234.
- Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology*, 27(4), 551-572.
- Winne, P. H. (2017). Cognition and metacognition within self-regulated learning. In *Handbook of self-regulation of learning and performance* (pp. 52-64). Routledge.
- Zhou, M., & Winne, P. H. (2012). Modeling academic achievement by self-reported versus traced goal orientation. *Learning and Instruction*, 22(6), 413-419.

AI Driven Support System (ADSS) for Teachers to Increase Underrepresented Minorities Academic Success

Taylor Stevens

George Mason University

tsteven8@gmu.edu

ABSTRACT: This work proposes a teacher-facing AI Driven Support System (ADSS) designed to assist in understanding and improving student performance, with a special focus on students from underrepresented groups. The need for this system stems from research that has shown underrepresented minorities continue to higher education and graduate at lesser rates than their non-minority counterparts. To help understand why this achievement gap exists it is imperative to look at a student's primary and secondary education, specifically focusing on the teachers. Researchers have found that a teacher's classroom practices correlate to student's academic performance. Additionally, they have found that underrepresented students perform better when they have a teacher of the same ethnic background, however, introducing more diversity between teachers can be an extensive task. Thus, we present an ADSS to help in closing the achievement gap between underrepresented students.

Keywords: Educational Data Mining (EDM), Learning Analytics, Underrepresented Minorities

1 INTRODUCTION

An underrepresented minority includes: Hispanic/Latinos, African Americans, Native Americans, Native Hawaiian/Pacific Islanders, and those of two or more races. For our purposes, due to lack of data for other racial groups, we will define African Americans and Hispanic/Latinos to be underrepresented minorities. There is a significant difference in the percent of underrepresented minorities that have a postsecondary education compared to their white counterparts. While many underrepresented minorities continue to a postsecondary education after high school only a small fraction of them graduate. For example, in 2014, 68.8% white students, 56.7% African American students, and 59.8% Hispanic students enrolled in college after completing high school (Digest of Education Statistics, 2018b). However, after 4 years, only 10% of African American students and 13% of Hispanic students graduated with a Bachelor's degree compared with the 61% of white students that did (Digest of Education Statistics, 2018a). These percentages follow the same trend for other years and degree levels.

It is estimated that by next year, about 2/3 of jobs will require a postsecondary degree (U.S. Department of Education, 2016a). Whether a student attends and graduates from a postsecondary institution is greatly influenced by their primary and secondary education. It was found that 56% of African American and 45% of Latino students were enrolled in remedial courses compared with 35% of white students (Jimenez, Sargrad, Morales, & Thompson, 2016). A remedial course is typically in Math or Reading/English to help improve the skills needed to succeed in the standard class. Students who take remedial classes have a greater chance of dropping out of college due to the increased degree completion time students (Jimenez, Sargrad, Morales, & Thompson, 2016). A student having to take a remedial course(s) can be indicative of

the education they received before attending college. Thus, in this study we plan to further explore the influences in an underrepresented student's academic performance.

There are many contributing factors when it comes to equity in education, with one of the most influential ones being teachers. One specific area to look into is the ethnicity of teachers. Research has shown that underrepresented minority students having a teacher of the same background can help close the achievement gap by teachers holding them to higher expectations and improving their overall school experiences (U.S. Department of Education, 2016b). However, a study at the Center for American Progress shows that in every state there is a higher percentage of students of color than teachers of color (U.S. Department of Education, 2016b). With the lack of diversity in teachers, some underrepresented students fail to get the attention they need to help them reach their maximum potential and succeed in school. Consequently, we arrive at the problem of underrepresented minority students either not continuing on to higher education or not completing their degree.

2 RELATED WORK

One of the most prevalent solutions to helping increase the diversity of teachers is financial and career/academic support in helping underrepresented minorities obtain their teaching certification. Another solution being discussed to increase the diversity amongst teachers is to create HR policies and practices that help to promote the hiring and retention of a diverse teaching staff (Motamedi & Stevens, 2019). Conversely, there are some solutions that aim to help the existing teachers. One proposed solution to help all teachers interact with different ethnic groups is continued education/professional development for diversity (Banks et al., 2001). In addition, it is suggested that teachers use a range of assessment strategies to help students from various backgrounds (Banks et al., 2001). Because these above solutions will likely require a great deal of funding and/or resources, we would like to look further and see if in general, regardless of race, whether a teacher's teaching style and resources influence an underrepresented student's performance.

There are already some studies that look at the relationship between the teacher and a student's performance, without considering either's race. A study done by Wenglinsky looked to evaluate a student's academic performance in relation to the teacher's quality (2002). Wenglinsky defines teacher quality to include the following: "the teacher's classroom practices, the professional development the teacher receives in support of these practices, and characteristics of the teacher external to the classroom, such as educational attainment" (2002). The study confirmed that a teacher's classroom practices had a significant effect on the 8th grade student's mathematics performance. In particular, it was found that professional development was the biggest influencer in a teacher's classroom practices (Wenglinsky, 2002). In sum, this study showed that for 8th grade students, their teacher's teaching style play a larger role than previously perceived in helping increase their academic performance in mathematics. Another study done in 2019 looked at the same relationship except using a different standardized mathematics test for 8th grade students. Similarly, the study found that there is a high correlation between a teacher's quality and a student's test scores (Burroughs et al., 2019). In addition to these studies, there are other ones that show the same association, but with varying test subjects and age groups. Hence, we intend to

explore this correlation on a wider scale so that we can get a better understanding of where the teachers could provide further guidance.

To help address these above issues, there has been increasing work looking into if and how AI applications can assist teachers. For example, Murphy looks into intelligent tutoring systems (ITS), automated essay scoring, and early warning systems, and how those assist teachers in the classroom (2019). ITS are geared more towards students while teachers monitor them to be able to give students the appropriate support. Some issues can arise with them because they are constrained to certain subjects and cannot assess higher-order skills. Additionally, it is hard for teachers to align the ITS's content with their instruction (Murphy, 2019). The second application, early warning systems, detect student's that are at risk of not graduating however, the system does not give concrete solutions for helping these students get back on track with graduation, it is up to teachers or administrators to do this. Thus, we would like to develop an easy to use system for all teachers that provides them with ample resources and guidance in assisting underrepresented minorities.

3 RESEARCH PROBLEM AND PROPOSED STUDY

Having an equivalent number of underrepresented teachers and students can improve student's performance however, it can be a difficult, costly, and lengthy task to introduce more diversity, amongst teachers, into schools. We have also seen that there are some AI systems that assist teachers, but these have some limitations and are for students of all racial backgrounds. So, in this study we aim to equip all the current teachers with the knowledge and resources to help underrepresented students. To do this, we want to create an AI Driven Support System (ADSS) for teachers to better understand an underrepresented student's general performance and how to increase their academic success.

Towards creating this system, we will address the following research questions:

1. What are the factors related to ethnicity that affects a student's performance on a test?
2. How does the ADSS need to be structured for it be useful for teachers?
3. What are the data requirements necessary for creating the ADSS?
4. What is the effectiveness of the ADSS and how will it be measured?

As a first step, we want to see how a student's performance correlates with whether they are an underrepresented minority. For this we will look at data provided by Educational Testing Service. The dataset contains the actions of 8th grade students through a mathematics National Assessment of Educational Progress (NAEP) assessment from the 2016-2017 school year. From this data, we look to train a model to predict a student's performance for the next section of the test. We would also like to understand the actions that were effective in helping them succeed in the next section and those that were not so that we can see if there is a pattern in where their skills lack. In addition to this dataset, data from the DataShop and ASSISTments which includes learning interaction data, will be used (Heffernan, 2014; Psicdatashop.web.cmu.edu, 2019). Similar methodology will be employed to these datasets so that a complete picture and training set of how students can best be assisted can be formed. Additionally, the mistakes made by the students, collected from these datasets, will be used in the ADSS to train it to make better suggestions to teachers.

To understand how the ADSS needs to be structured, it is necessary to conduct some structured interviews and develop a survey for teachers to understand their needs and wants for the proposed system. The survey will include questions regarding a teacher's teaching style and the average performance of their students. This will allow us to see if there is a correlation between a teacher's style and student's performance. To get a better understanding of the style of the ADSS that will be most effective for teachers, we will give them a few proposed prototype designs of the system and see which they feel would be the most beneficial and why. Additionally, the survey will include general questions on recommender systems. For example, if they already use a recommender system, which one do they use and what do they like/dislike about it. The results from this question as well as the first research question will help in deciding which components should go in the system.

After establishing the structure and components of the system, it is very important to understand the input format and structure of the data in the ADSS. Some things we currently know that the data will need to include is the type of question (multiple choice, short answer, etc.), the topic of the question, the answer input and correct answer, and where the student made a mistake (if applicable). Additionally, from the second research question we will understand what other attributes the data needs to include in order to incorporate what the teachers identified as their needs. Lastly, for evaluating the effectiveness and efficiency of the system, we need to answer common questions asked when assessing recommender systems. We will try to address the general goals of recommender systems given by Aggarwal in *Recommender Systems*. These include questions on accuracy, coverage, trust, novelty, diversity, and scalability (Aggarwal, 2016). Additionally, another survey asking teachers questions regarding their use of the ADSS and the underrepresented student's performance may be useful in understanding the system's effectiveness.

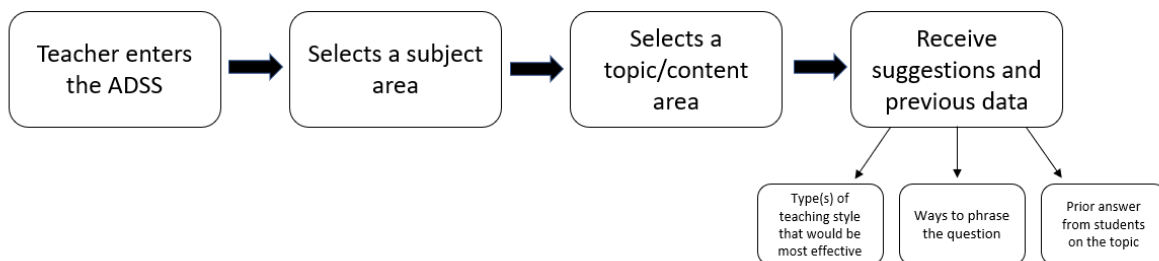


Figure 1: General concept of the ADSS

Figure 1 gives the general concept of the ADSS. A possible interaction the teacher could have with this system is as follows. Suppose a teacher is having difficulty explaining ratio word problems. They would enter the system, select “Math”, then “Ratios” and a list of different ratio problems would be displayed. The teacher would then select a ratio word problem and student’s responses and common mistakes would be returned. The teacher could then select a common problem and a recommendation would be made available. For example, if one of the common problems was that a student did not understand the context

and wording of the problem, the teacher may receive alternative wording used by other teachers that was found effective.

While this type of system is not ideal for every subject, through further research we will find out what subjects it would be most effective for. Furthermore, while the focus of this system, as of now, is for underrepresented minorities, it can be expanded to benefit other groups such as first-generation students.

4 CURRENT STATUS OF THE WORK

Currently, the mathematics NAEP assessment dataset is being preprocessed, using R, to represent each question a student answers as one instance. Also, review into the NAEP tests, specifically the questions and what content areas/skills the tests measure is being done, as well as collection and analysis of the DataShop and ASSISTments data.

REFERENCES

- Aggarwal, C. (2016). *Recommender Systems*. Switzerland: Springer, pp.225-252.
- Banks, J. A., Cookson, P., Gay, G., Hawley, W. D., Irvine, J. J., Nieto, S., ... Stephan, W. G. (2001). Diversity within Unity: Essential Principles for Teaching and Learning in a Multicultural Society. *Phi Delta Kappan*, 83(3), 196–203. doi: 10.1177/003172170108300309
- Burroughs N. et al. (2019) Relationships Between Instructional Alignment, Time, Instructional Quality, Teacher Quality, and Student Mathematics Achievement. In: Teaching for Excellence and Equity. IEA Research for Education (A Series of In-depth Analyses Based on Data of the International Association for the Evaluation of Educational Achievement (IEA)), vol 6. Springer, Cham
- Digest of Education Statistics. (2018a). Bachelor's degrees conferred by postsecondary institutions, by race/ethnicity and sex of student: Selected years, 1976-77 through 2016-17. Retrieved from https://nces.ed.gov/programs/digest/d18/tables/dt18_322.20.asp.
- Digest of Education Statistics. (2018b). Percentage of recent high school completers enrolled in college, by race/ethnicity: 1960 through 2017. Retrieved from https://nces.ed.gov/programs/digest/d18/tables/dt18_302.20.asp.
- Heffernan, N. (2014). *ASSISTmentsData*. [online] ASSISTmentsData. Available at: <https://sites.google.com/site/assistmentsdata/>.
- Jimenez, L., Sargrad, S., Morales, J., & Thompson, M. (2016, September 28). Remedial Education. Retrieved from <https://www.americanprogress.org/issues/education-k-12/reports/2016/09/28/144000/remedial-education/>.
- Motamedi, J. G., & Stevens, J. (2019, February 8). Increasing Diversity in the Teacher Workforce: The Importance and Potential Impact of Authentic Change. Retrieved from <https://ies.ed.gov/ncee/edlabs/regions/northwest/blog/increasing-diversity.asp>.
- Murphy, R. (2019, January 23). Artificial Intelligence Applications to Support Teachers. Retrieved from <https://www.rand.org/pubs/perspectives/PE315.html>.
- National Center for Education Statistics. (2019, July 18). Questionnaires for Students, Teachers, and

- School Administrators. Retrieved from https://nces.ed.gov/nationsreportcard/experience/survey_questionnaires.aspx.
- Pslcdatashop.web.cmu.edu. (2019). *DataShop@CMU*. [online] Available at: <https://pslcdatashop.web.cmu.edu/>.
- U.S. Department of Education. (2016a). *Advancing Diversity and Inclusion in Higher Education* [PDF file]. Retrieved from <https://www2.ed.gov/rschstat/research/pubs/advancing-diversity-inclusion.pdf>.
- U.S. Department of Education. (2016b). *The State of Racial Diversity in the Educator Workforce* [PDF file]. Retrieved from <https://www2.ed.gov/rschstat/eval/highered/racial-diversity/state-racial-diversity-workforce.pdf>.
- Wenglinsky, H. (2002, February 13). How schools matter: The link between teacher classroom practices and student academic performance. *Education Policy Analysis Archives*, 10(12). Retrieved from <http://epaa.asu.edu/epaa/v10n12/>.

The 6th LAKathon: Accelerating Development by Learning from the Past

Daniele Di Mitri

Open University of the
Netherlands
daniele.dimitri@ou.nl

Gábor Kismihók

TIB Hannover
gabor.kismihok@tib.eu

Alan Mark Berg

ICTS, University of Amsterdam
a.m.berg@uva.nl

Kirsty Kitto

University of Technology,
Sydney
kirsty.kitto@uts.edu.au

Stefan T. Mol

University of Amsterdam
s.t.mol@uva.nl

Jan Schneider

DIPF - Frankfurt
schneider.jan@dipf.de

Jose Ruipérez-Valiente

University of Murcia,
Information and
Communications Engineering
jruipeerez@um.es

ABSTRACT: Welcome to the sixth Learning Analytics Hackathon (LAKathon). This year, the LAK conference celebrates its tenth anniversary. But what will the Learning Analytics community look like ten years from now? The LAKathon 2020 will become the laboratory to envisage future Learning Analytics (LA) applications. Do you have a research question, a dataset or a futuristic idea you would like to explore? Bring it to the LAKathon! We encourage joining this inclusive workshop no matter what your background or skills, everyone is welcome. We aim to address the science-practice divide by having practitioners and researchers from diverse fields working in multidisciplinary teams towards common objectives.

Keywords: Hackathon, Future, Learning Analytics, Multimodal, Interoperability, xAPI

1 INTRODUCTION

For the last five years, researchers and practitioners have run hackathons at the Learning Analytics & Knowledge (LAK) conferences. We have brainstormed on new LA techniques, discussed technical infrastructures and analysed educational datasets. We have formed opinions and suggested strategies which have radiated back to the LA research community as a whole. In 2020, in the 10th anniversary of the LAK conference, the theme chosen for the 6th LAKathon is *“Accelerating Development by Learning from the Past”*. In this edition, we suggest key thematic objectives which build upon the key topics of previous’ LAKathon editions. All LAKathons have been designed to be 1) *solution-driven*, participants solve a series of realistic challenges through *Agile* research approaches, including brainstorming, design thinking, or fast-prototyping; 2) *multi-disciplinary*, reflected in the diversity of the LAKathon participants; and 3) *self-organised*, as we engage in bottom-up, and actionable research questions. Through a Call for Proposals, we aim to elicit research questions that address the thematic objectives thus accelerating *the development* of the LAK community. We aim also to strengthen the bridge between the LAKathon and the LAK conference by explicitly inviting the LAK research sub-communities and Special Interest Groups to join the LAKathon and propose their challenges, as it was the case last year for the CrossMMLA SIG. The LAKathon 2020 wants to become the space for hands-

on technical challenges, which take place in parallel to the work of the LAK sub-communities and offers them the space to address challenges.

2 BACKGROUND

LAKathon 2015 focused on the *Apereo Open Dashboard*, with data sourced from an Experience API (xAPI) Learning Record Store (LRS). It illustrated how the concept of an Open Learning Analytics architecture can be made and discussed what a learning analytics dashboard must contain.

LAKathon 2016 explored *Open Learning Analytics*. Using as reference point the emerging Learning Analytics architecture developed by Jisc (Sclater, Berg, & Webb, 2015). The participants scrutinised Jisc's interoperability recipes, tested the interoperability of LRS, and Dashboards and sought to work towards unifying the standards landscape for learning analytics. The hackathon had a lasting effect, with numerous improvements to Jisc's interoperability recipes. A strong message from the LAK community, known as the *Edinburgh Statement*, set the basis for greater integration of two emerging learning analytics standards: Experience API and IMS Caliper (Apereo, 2018).

LAKathon 2017 built upon three assets: previous workshops, research, and recently-developed software. The first comprised the previous two LAKathons, and two previous workshops "Visual Aspects of Learning Analytics" and "Data Literacy for Learning Analytics". The second involved recent research on actionable analytics, student feedback, and embedding learning analytics in pedagogic practice (Kitto, et al. 2016). The third, finally, involved the introduction of Jisc's student app, which was piloted with students across the UK after extensive consultation and design activities.

LAKathon 2018 saw a continuation, expansion and documentation of previous themes. The challenges were goal setting for portfolios and employability, sensor-based and multimodal learning analytics, and the creation of Data Literacy Playground. The LAKathon 2018 also looked into algorithmic transparency and ethical workflows.

LAKathon 2019 revolved around three main challenges: the Interoperability Challenge, sought synergies between xAPI and Tin Can API profiles. The Game-based analytics challenge, which aimed at creating a process to integrate LA in game-based assessment (Kim et al. 2019) and wondered how to detect when students are stuck and disengaged. The third challenge which envisioned a markup language to describe blended learning courses was curriculum analytics. This LAKathon challenge created a JSON markup which can qualify a curriculum both in distance and in lab learning scenarios.

3 OBJECTIVES

The expected outcomes of the LAKathon are the identification and concrete pilot implementations of prototype tools/systems/data/studies, which arise from the synthesis of educational technology, software development, and data science perspectives. As for previous events, the hackathon will generate a repository of code, sample data, screenshots, and slides from the activity of participants. At LAKathon2020 we expect to emphasise the following topics

Multimodal Learning Analytics. Learning activities such as practical skills training and co-located group interactions represent a big set of learning moments taking place across physical and digital

spaces, both in the classroom and at the workplace. These moments can be monitored via the tracking of multiple modalities including motoric and physiological information, learning context, environment, and activity (Di Mitri et al., 2018). This year the MMLAHack seeks to complement the conference workshop CrossMMLA with a space for hands-on examples, prototype demonstrations, code-sharing, and solutions to technical issues in the field of MMLA.

Data Interoperability. It should not matter which institution a student attended, their learning data should make sense for a lifetime and in any environment. This becomes even more urgent in light of GDPR legislation and the concomitant right to data portability. In LAK20 we propose to return to the Edinburgh Statement on Data Interoperability (Edinburgh Statement, 2016), exploring ways of mapping between emerging xAPI Profiles, and published Caliper metric profiles. Bring your xAPI and Caliper data along and help us to find a way to develop LA tools that are agnostic about which data standard was used to generate that data.

Goal setting and analytics: Goal Setting (GS) theory, and GS tools and methodologies, can potentially enhance the performance of individuals. Tracking students' learning through GS comes with a number of opportunities to gain insights into learning pathways, Self Regulated Learning and offline learning activities that students engage in. With the help of learning analytics, it is possible to connect students' learning goals with performance and behavioural data coming from digital learning environments. This notion also creates the opportunity to continuously monitor students' progress toward their explicit learning goals over time, and provide individual recommendations. We will build on the LAK16 Goal setting workshop (Mol, Kobayashi, Kismihók & Zhao, 2016), available open-source applications (including the UvA Goal Setting Dashboard), and a large amount of labour market data (vacancy announcements) to formulate personal goals beyond the frames of formal education.

Analytics in game-based assessment and games for learning: Well designed games represent wonderful opportunities for learning and as assessment instruments for numerous reasons (Kim et al. 2019): They keep players engaged through the process, most people have grown up playing games and they relate them with fun, and games can represent more closely real world situations than traditional learning environments. As a challenge, due to the often open nature of this medium, it is complex to make sense of these kind of data. This challenge explores the affordances of using analytics to understand students' learning process and to perform assessment in games.

Curriculum Analytics How can we enable the identification of effective or weak parts of the curriculum, leading to the building of better courses? Early efforts at linking learning design and learning analytics include the Australian "Loop" system, which integrates course structures and schedules in its visualisations to help evaluate the effectiveness of the learning activities and work at the Open University which assesses the impact of types of learning and assessment design on various measures of student success (Quan Nguyen, 2017). An approach we intend to explore is the building of learning designs which not only categorise different aspects of the learning process but also specify the data which needs to be captured to show whether the designs are proving effective.

4 ORGANISATIONAL DETAILS

The LAKathon is organised as an open workshop over two days. The first half-day is a period for participants to get acquainted with the core themes such as how to interact with the data and the

tools. We then divide into teams of 6-8 people to fulfil specific missions. At the end of the first day we discuss progress, lessons learnt and next steps. At the end of the second day, we summarise and plan future follow-up actions. To strengthen the linkage of the LAKathon with other communities at LAK, this year we want to allow one-day participation to the LAKathon. We also propose a Call for Proposals to the whole LA community for short and on-point submissions (1 page long) detailing research questions and associated datasets. For the logistics, we need a large room (50 participants) with a reliable internet connection, projector, separate tables for group exercises, and if possible stationery such as sticky notes and pens. The organisers will provide technical tools and LA infrastructures, seed datasets, Git repository, Slack channel, and disseminate both progress and outcomes via blogs and the Twitter hashtag: #LAKathon.

REFERENCES

- Apereo (2018). Apereo Open Dashboard. Retrieved from <https://apereo-learning-analytics-initiative.github.io/OpenDashboard/>
- Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning*, 34(4), 338-349.
- Duval, E. et al. (2015) VISLA 2015, Visual Aspects of Learning Analytics. *CEUR Workshop Proceedings*.
- Edinburgh Statement for Learning Analytics Interoperability (2016)
<https://docs.google.com/document/d/18VJ9hVcfk9sOzs2MEeZ59Q5RSz1jZnUHhSWFSednEXg/edit#heading=h.votza0o0t5xa>
- Kitto, K. et al. (2016) Incorporating student-facing learning analytics into pedagogical practice. *Proceedings of the Annual ASCILITE Conference* (2016).
- Mol, S., Kobayashi, V., Kismihók, G., & Zhao, C. (2016). Learning Through Goal Setting. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 512–513). New York, NY, USA: ACM. <https://doi.org/10.1145/2883851.2883859>
- Quan Nguyen, Bart Rienties, Lisette Toetenel, Rebecca Ferguson, Denise Whitelock, Examining the designs of computer-based assessment and its impact on student engagement, satisfaction, and pass rates, *Computers in Human Behavior* 76 (2017) 703-714
- Sclater, N., Berg, A., & Webb, M. (2015). Developing an open architecture for learning analytics. *Proceedings of the EUNIS 2015 Congress*. <https://doi.org/ISSN>. pp. 2409–1340.
- Kim Y.J., Ruipérez-Valiente, J. A. Tan P., Rosenheck L. and Klopfer E. Towards a Process to Integrate Learning Analytics and Evidence-Centered Design for Game-based Assessment. *Companion Proceedings of the 9th International Learning Analytics and Knowledge Conference (LAK'19)*, (pp. 204-205). Tempe (AZ), USA. 2019.

Linking Individual Contribution Assessment in Collaborative Learning with Learning Analytics: A Multimodal Approach

Song Lai, Hao Tian, Jiaqi Liu and Fati Wu

Beijing Normal University

{laisong, tianhao}@mail.bnu.edu.cn, liujiaqi1212@foxmail.com, wft@bnu.edu.cn

ABSTRACT: In collaborative learning, individual contributions have a direct effect on group performance, so their accurate assessment is crucial. Educational data mining techniques can extract useful information from multimodal data to assess individual contributions during the process and of the products in collaborative learning. In this LAK Hackathon proposal, the current study presents a challenge to build an individual contribution assessment framework using learning analytics techniques in the collaborative learning environment.

Keywords: Learning analytics, multimodal data, individual contribution assessment

1 BACKGROUND

In recent years, the benefits of collaborative learning have been reported widely in numerous empirical studies (Appavoo, Sukon, Gokhool, & Gooria, 2019). Collaborative learning is an educational practice where two or more students perform a considerable part of the task. Individual efforts of group members need to be acknowledged (Johnston & Miles, 2004; Le, Janssen, & Wubbels, 2017; Xu, Zheng, Hu, & Li, 2016). The lack of individual efforts may result in social loafing and free-riding (Khandaker & Soh, 2010). The two phenomena usually lead to a low contribution to group product. In order to score objectively group members, assessing accurately individual contributions during a collaborative process is crucial (Zhang & Ohland, 2009). Once realizing that contributions can be reasonably rewarded, students' motivation, perception, and participation in over group work could be further improved. Subsequently, the collaboration may be enforced and above phenomena could be avoided, leading to the improvement of individual contributions and even group performance. Hence, the analysis of individual contributions can more exactly rate students in collaborative learning. Also, the information referring to the distribution of individual contributions in over group work can support instructors to provide appropriate guidance.

In a collaborative learning environment, it is usually difficult for instructors to monitor collaborative processes and individual contributions (Le et al., 2017). With the help of learning analytics techniques, individual contributions can be assessed by extracting and analyzing useful information from collaborative data. At present, most of the existing studies apply quantitative analysis (Ding, 2009), social network analysis (Li, Liao, Wang, & Huang, 2007), content analysis (Daradoumis, Martínez-Monés, & Xhafa, 2006), and self- and peer-assessment (Ma, Yan, & Wang, 2018) methods to analyze individual contributions. These methods have demonstrated the effective ability of individual contribution assessment. However, notably few studies have been conducted to automatically assess individual contributions by combining multimodal data, such as online operation interaction, text, audio, video, social network and product data. Hence, this study will

present a challenge about the construction of a mixed framework to measure individual contributions from multimodal data in collaborative learning.

2 RESEARCH QUESTIONS

- 1) How to extract multimodal data in the collaborative learning process?
- 2) Which data are valuable or meaningful to individual contribution assessment?
- 3) Which learning analytics techniques form the best model assessing individual contributions?

3 EXPECTED OUTCOMES

This study extracts multimodal collaborative data using effective educational data mining methods. The collected data can be analyzed to identify valuable or meaningful data which contribute to individual contribution assessment. Following this, a mixed framework to measure individual contributions can be constructed. Finally, the study will find out a best learning analytics technique to assess individual contributions. In order to deliver real-time information feedback for instructors, the automated quantitative results about individual contributions will be visualized using graphs.

REFERENCES

- Appavoo, P., Sukon, K. S., Gokhool, A., & Gooria, V. (2019). Why does collaborative learning not always work even when the appropriate tools are available? *Turkish Online Journal of Distance Education*, 11-30.
- Daradoumis, T., Martínez-Monés, A., & Xhafa, F. (2006). A layered framework for evaluating online collaborative learning interactions. *International Journal of HC Studies*, 64, 622-635.
- Ding, N. (2009). Visualizing the sequential process of knowledge elaboration in computer-supported collaborative problem solving. *Computers & Education*, 52, 509-519.
- Johnston, L., & Miles, L. (2004). Assessing contributions to group assignments. *Assessment and Evaluation in Higher Education*, 29, 751-768.
- Khandaker, N., & Soh, L. K. (2010). Assessing individual contributions to groupwork by a summary of tracked computer-supported collaborative activities. *2010 IEEE Frontiers in Education Conference* (pp. 1-6).
- Le, H., Janssen, J., & Wubbels, T. (2017). Collaborative learning practices: Teacher and student perceived obstacles to effective student collaboration. *Cambridge Journal of Education*, 1-20.
- Li, Y., Liao, J., Wang, J., & Huang, R. (2007). CSCL interaction analysis for assessing knowledge building outcomes: method and tool. in *Proceedings of the 8th international conference on computer supported collaborative learning* (pp. 431-440).
- Ma, Z., Yan, X., & Wang, Q. (2018). Assessing individual contribution in collaborative learning through self- and peer-assessment in the context of china. *Innovations in Education & Teaching International*, 1-12.
- Xu, C., Zheng, Y., Hu, H., & Li, Y. (2016). Exploring individual contributions during online collaborative discussions. *Journal of Computers in Education*, 3, 395-411.
- Zhang, B., & Ohland, M. W. (2009). How to assign individualized scores on a group project: an empirical evaluation. *Applied Measurement in Education*, 22(3), 290-308.

Assessing Learning in Real Time – Learning Analytics strategy for classrooms

Vijayan N

Temasek Polytechnic, Singapore

vijayan@tp.edu.sg

ABSTRACT: This proposal sets out a research question for the 6th Learning Analytics Hackathon @LAK 2020. The focus is on exploring methods and techniques to garner real time feedback on learning in the classroom from hackathon participants. This is for a live project, Assessing Learning in Real Time (ALERT), that allows for real time feedback from learners to be obtained via student mobile devices at the end of a lesson. This feedback was visualised in a dashboard for the tutor to assess learning issues faced by students. This allowed for tutors to redesign their lessons for the subsequent lessons and address the concerns raised by learners.

Keywords: Feedback on Learning; Real Time Feedback; Diagnostic Assessment

1 BACKGROUND

This hackathon proposal builds on past presentations at LAK conferences (especially 2018 & 2019) in the use of real time feedback on students' learning (Mori, Sakamoto and Mendori, 2019; Chen, Saleh, Hmelo-Silver, Glazewski & Lester, 2019; Azcona, Hsiao & Smeaton, 2018; Moxley & Bennington, 2018; Matcha, 2018).

The use of analytics in education is gaining momentum. The use of access data from LMS and other sources that students use as an indication of their preparedness and learning is providing useful learning analytics to teachers. However, as the LOOP project (Corin, et al. 2016) states in the discussion segment of the project implementation, the project's intent to "identify, diagnose and resolve issues with learning" (p.43) was not adequately addressed during project implementation. The authors rightly pointed out that the learning analytics should be utilized beyond access data to address learning and teaching challenges faced by tutors. Similarly, Long and Siemens (2011) had also postulated that the field of learning analytics provides the basic of capturing feedback from learners on the learning, teaching and environment. Learning analytics provides a lens through which tutors are able to diagnose learning occurrences for the purpose of analysis. Such a focused approach allows for educators to bring learning analytics instruments into the ambit of their teaching and learning environment, for a variety of purposes.

ALERT – Assessing Learning in Real Time, is a project implemented for this very purpose. The relevance of getting real time feedback from learners and adapting lessons to meet the needs of learners is the focus of this implementation at various Institutes of Higher Learning (IHL) in Singapore.

2 ASSESSING LEARNING IN REAL TIME (ALERT)

ALERT was conceptualized to provide real time data to better understand learners and their needs at the end of a learning task or activity. The data captured acts as feedback from learners, regarding the learning activity, and allows for tutors to design appropriate intervention activities to address the concerns during the subsequent lesson. Such real time feedback and intervention is done to

directly support and enhance the learning progress of learners more efficiently. The evolution of learning analytics (LA) systems and tools offers vast opportunities for educators to harness the process of analytics for diagnostic purposes in the learning and teaching context.

3 RESEARCH QUESTION

The proposed research question for the hackathon @LAK2020 is then, “What other methods and techniques can be used to garner real time feedback on learning in the classroom?”

Two techniques to gather students learning analytics have been deployed. The first requires asking students to provide feedback on their learning at the end of the lesson using a Likert scale or open-ended response. The second involves designing a diagnostic question on the lesson done. The answers to the diagnostic question would then reflect the students’ learning to tutors.

Apart from these two methods, the researcher is exploring other techniques and/or methods to gather evidence on students’ learning in real time. By tapping on the expertise of participants at LAKathon 2020, more methods and techniques could be explored, and assimilated into ALERT.

4 EXPECTED OUTCOMES

The practical solutions garnered from the 6th Learning Analytics Hackathon will be presented to the Learning Analytics sub-committee and scheduled for implementation across the IHLs in Singapore. The implementation effect will be studied for enhancement to be done for the ALERT project.

REFERENCES

- Azcona, Hsiao & Smeaton, 2018; PredictCS: Personalizing Programming Learning by Leveraging Learning Analytics. In Companion Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK'18) (pp464-470).
- Chen, Y., Saleh, A., Hmelo-Silver, C., Glazewski, K, & Lester, J. (2019). Preparing Successful Facilitation: Designing A Teacher Dashboard to Support PBL Classroom Orchestration in A Game-based Learning Environment. In Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge (LAK'19) (pp.218-220).
- Corrin, L., Kennedy, G., de Barba, P.G., Lockyer, L., Gašević, D., Williams, D., Dawson, S., Mulder, R., Copeland, S., & Bakharia, A. (2016). Completing the Loop: Returning Meaningful Learning Analytic Data to Teachers. Sydney: Office for Learning and Teaching.
- Long, P., & Siemens, G. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*, 46, 30–32. <https://doi.org/10.1145/2330601.2330605>.
- Matcha, W. (2018). Beyond Data Presentation: Learning Analytics to Uncover Learning Strategies and the Influences of Feedback. In Companion Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK'18) (pp211-218).
- Mori, Y., Sakamoto, K., & Mendori, B. (2019). Development of a Real Time Viewing Status Feedback System and Its Impact. In Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge (LAK'19) (pp.174-175).
- Moxley, J. & Bennington, S.T. (2018). Actionable Analytics at MyReviewers for Administrators, Instructors, Students, and Researchers. In Companion Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK'18) (pp535-544).

Learning Analytics for Classroom Activities

Jan Schneider

DIPF

Schneider.jan@dipf.de

Daniele Di Mitri

OUNL

Daniele.Dimitri@ou.nl

George-Petru Ciordas-Hertel

DIPF

ciordas@dipf.de

ABSTRACT: Learning Analytics (LA) mostly focuses on tracking and analyzing interactions between learners and online learning environments. Learning, however, happens in a physical environment and in many cases it happens without the direct interaction with online environments. When the physical environment is not aligned with the learning tasks the quality of education suffers. In this project, we aim to address how LA can be used to support the learning that happens in a classroom.

Keywords: LA for Classrooms, Smart Learning Environments, Multimodal Learning Analytics

1 BACKGROUND

The physical space has a big influence on learning. Imagine studying from a textbook in a dark room, solving a math problem in the middle of a rock concert or learning to swim following an online course while sitting in front of your laptop. When pedagogies are not aligned with the physical spaces, the quality of education suffers (van Merriënboer, McKenney, Cullinan & Heuer, 2017). In this project, we will investigate how Learning Analytics can be applied to enhance learning in physical spaces, more specifically in a classroom.

1.1 Research Questions and Expected outcomes

- What effective learning activities are in the classroom?
- How can we collect relevant information about these activities?
- How can we adapt the environment based on the learning activities?

The expected outcome is to design a wizard that based on the learning tasks it informs what adaptations to the environment should be made and what needs to be tracked.

REFERENCES

van Merriënboer, J. J. G., McKenney, S., Cullinan, D., & Heuer, J. (2017). Aligning pedagogy with physical learning spaces. *European Journal of Education*, 52(3), 253–267.

Applicability of Automatic Short Answer Grading Systems in Assessment Scenarios

Anna Filighera

KOM - TU Darmstadt

anna.filighera@kom.tu-darmstadt.de

Tim Steuer

KOM- TU Darmstadt

tim.steuer@kom.tu-darmstadt.de

ABSTRACT: The goal of this LAKathon project is to investigate the strengths and weaknesses of the state of the art automatic short answer grading system based on BERT. For this, participants of this project will be challenged to systematically outsmart the grading model.

Keywords: Assessment, Automatic Short Answer Grading, Adversarial Examples

1 BACKGROUND

Automatic short answer grading systems judge the factual correctness, completeness and relevance of student answers to a given question without human involvement. In contrast to automatic essay grading, only the semantics of the short answers are relevant for grading instead of also considering the writing style etc. While automatic short answer grading systems become more widespread their accuracy is still far from perfect. Our proposal is to do a systematic investigation of their strengths and weaknesses. For this we will be working with the BERT based state of the art approach proposed by Sung et al. (Sung et al., 2019) on the SemEval 2013 benchmark dataset (Dzikovska et al., 2013).

2 RESEARCH QUESTIONS

We aim to explore the following research questions during the LAKathon2020:

1. To what extent are automatic short answer grading systems vulnerable to manually crafted adversarial examples?
2. To what extent are automatic short answer grading systems robust with regard to natural occurring answer variations?
3. To what extent can automatic short answer grading systems complement human grading?

3 EXPECTED OUTCOMES

We expect the research questions to be investigated in the following steps:

Regarding RQ 1 we encourage the participants to attack given state of the art automatic short answer grading models. The first result of this workshop consists of the concrete, manually crafted adversarial examples. Based on these individual cases generalized attack patterns could emerge.

To answer RQ 2 an analysis of naturally occurring variations of answers will be conducted and tested on the model. Such variations may include paraphrasing or more or less detailed elaboration.

On the basis of the results of RQ 1 and 2, possible applications of automatic short answer grading systems are discussed and documented.

REFERENCES

- Dzikovska, M.O., Nielsen, R.D., Brew, C., Leacock, C., Giampiccolo, D., Bentivogli, L., Clark, P., Dagan, I., Dang, H.T.: Semeval-2013 task 7: The joint student response analysis and 8th recognizing textual entailment challenge. Tech. rep., NORTH TEXAS STATE UNIV DENTON (2013)
- Sung, C., Dhamecha, T.I., Mukhi, N.: Improving short answer grading using transformer-based pre-training. In: International Conference on Artificial Intelligence in Education. pp. 469–481. Springer (2019)

Benchmarking Ethical LA algorithms

Alan Mark Berg

ICTS, University of Amsterdam
a.m.berg@uva.nl

Shereif Eid

TIB Hannover
Shereif.Eid@tib.eu

Gábor Kismihók

TIB Hannover
gabor.kismihok@tib.eu

Stefan T. Mol

University of Amsterdam
s.t.mol@uva.nl

ABSTRACT: There is an ongoing debate (Mittelstadt et al, 2016; Jobin, Ienca & Vayena, 2019) on the ethical use of AI and Big data. Consider for example the increased usage in Learning Management Systems and MOOCs. Tools within the remit of Learning Analytics deploy AI at scale. Therefore, it is inevitable that AI based decisions will become more embedded and impactful. During this Lakathon, the authors seek to prototype an initial benchmark that defines and compares the relative bias of algorithms based on the properties of an ethical algorithm.

Keywords: Ethical Algorithm, Library, AI Benchmark

1 BACKGROUND

There is an ongoing debate (Mittelstadt et al, 2016; Jobin, Ienca & Vayena, 2019) on the ethical use of AI and Big data. There are also benchmarks for discriminatory bias based on an ensemble of metrics (Bellamy et al, 2018). Within the field of Learning Analytics there is a need for a universally accepted ethically orientated set of benchmarks for Learning Analytics related algorithms. The lack of such benchmarks perpetuates the deployment of sub optimal algorithms within Educational software such as Learning Management Systems and MOOCs. Due to the lack of universal validation practices it is difficult for decision makers to choose the best products for students and teachers. During the Lakathon, the authors seek to prototype an initial benchmark that defines and compares the relative bias of algorithms based on the yet to be defined properties of what an ethical algorithm would like. Moreover, we seek to highlight the notion of trust with respect to LA algorithms. In this context, we will investigate the main characteristics of a trusted LA algorithm and review the potential of reliable trust metrics.

2 RESEARCH QUESTIONS

During the Lakathon we will address the following research questions:

RQ 1: *Within the context of LA what is an ethical algorithm?*

RQ 2: *What is 'Trust' in the context of LA algorithms?*

RQ 3: *What are the trust metrics that should ideally identify a LA algorithm ?*

RQ 4: *What are the properties, metrics, and practices that uniquely define an ethical LA algorithm?*

RQ 5: *How do we benchmark an ethical LA algorithm?*

RQ 6: *How do we simplify and standardize the collection of LA related data sets relevant to benchmarking?*

3 EXPECTED OUTCOMES

During the Lakathon we expect to generate a range of useful products and prototypes that may then be refined in further events. The initial planning is for the following products:

Product 1: An initial definition of the properties, metrics, and practices necessary to evaluate whether an algorithm may be ethically deployed within the context of Learning Analytics.

Product 2: A prototype library which makes visible well described LA data sets.

Product 3: A library that abstracts benchmarking for LA algorithms and connects to well described LA data sets

Product 4: A list of future research questions including motivations.

REFERENCES

- Bellamy, R. K., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., & Nagar, S. (2018). AI fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. *arXiv preprint arXiv:1810.01943*. Source code: <https://github.com/IBM/AIF360>
- Jobin, A., Ienca, M., & Vayena, E. (2019). Artificial Intelligence: the global landscape of ethics guidelines. *arXiv preprint arXiv:1906.11668*.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*. <https://doi.org/10.1177/2053951716679679>

Supporting Learning Design Activities Outside the Classroom with Learning Analytics

Gerti Pishtari, María Jesús Rodríguez-Triana

Tallinn University

gerti.pishtari@tlu.ee, mjrt@tlu.ee

ABSTRACT: In this LAK Hackathon proposal, we present the challenge of creating learning analytics that support the learning design practices of Smartzoos, a location-based authoring tool with game elements.

Keywords: Location-based learning, Game-based Learning, Learning Design, Learning Analytics, Visualisations

1 BACKGROUND

Mobile and ubiquitous technologies such as location-based tools can transform situated environments (like archeological parks, or museums) into learning scenarios. Despite the affordances of these technologies, designing, monitoring and evaluating learning is more complex since it takes place in a distributed environment. To overcome these challenges, there is a growing interest in aligning learning design practices with learning analytics (Hernández-Leo et. al, 2019; Lockyer, 2013). On the one hand, learning design can make the analysis more meaningful, while learning analytics can support evidence-based decisions about the design (Hernández-Leo et. al, 2019; Lockyer, 2013). Apart from practitioners, this alignment could be beneficial also for other stakeholders with an interest around a specific learning design tool (e.g., the community of practitioners or researchers around the tool) (Hernández-Leo et. al, 2019, Pishtari et. al, 2019).

We have deployed Smartzoos¹, a location-based authoring tool that supports the design of gamified learning scenarios outdoor (Pishtari et. al, 2017). It enables the creation of a set of geo-localised quizzes and puzzles, and is able to collect a set of multimodal data about students' interaction with the game and the surrounding environment. Currently, the main stakeholders dealing with the tool are practitioners who create content, zoo managers, researchers, and developers. As part of the LAK Hackathon, we will present a dataset from Smartzoos and a short demonstration of a playable scenario.

2 RESEARCH QUESTIONS

Looking at the Smartzoos dataset, the main research questions to be explored during the workshop are:

¹ <https://smartzoos.eu/>

1. How can learning analytics inform the learning design practices of the stakeholders of location-based authoring tools such as Smartzoos? What additional information would be necessary to collect?
2. What visualisations could be used to represent the data outcome?

3 EXPECTED OUTCOME

The expected outcome would be a set of algorithms and visualisations that would respond to the research questions. Moreover, the results obtained from the session could provide the basis for a future publication among the Smartzoos team and the participants of the LAK Hackathon that will work on this session.

ACKNOWLEDGMENTS

THIS RESEARCH HAS BEEN PARTIALLY FUNDED BY THE EUROPEAN UNION IN THE CONTEXT OF CEITER (HORIZON 2020 RESEARCH AND INNOVATION PROGRAMME, GRANT AGREEMENT NO. 669074) AND SMARTZOOS, CB64, CENTRAL BALTIC PROGRAMME 2014-2020. THE AUTHORS WOULD LIKE TO THANK THE TEAM OF INTERVIEWERS FROM TALLINN UNIVERSITY FOR THEIR CONTRIBUTION TO THIS STUDY.

REFERENCES

- Hernández-León, D., Martínez-Maldonado, R., Pardo, A., Muñoz-Cristóbal, J. A., & Rodríguez-Triana, M. J. (2019). Analytics for learning design: A layered framework and tools. *British Journal of Educational Technology*, 50(1), 139-152.
- Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist*, 57(10), 1439-1459.
- Pishtari, G., Rodríguez-Triana, M. J., & Väljataga, T. (2019, September). Multi-stakeholder Analytics for Learning Design: A Case Study of Location-based Tools. In *World Conference on Mobile and Contextual Learning* (pp. 94-101).
- Pishtari, G., Väljataga, T., Tammets, P., Savitski, P., Rodríguez-Triana, M. J., & Ley, T. (2017, September). SmartZoos: modular open educational resources for location-based games. In *European Conference on Technology Enhanced Learning* (pp. 513-516). Springer, Cham.

Data-driven Assessment of Spatial Reasoning in a Geometry Game through Learning Analytics

José A. Ruipérez Valiente

University of Murcia, Faculty of Computer Science
jruiperez@um.es

Yoon Jeon Kim

MIT, Playful Journey Lab
yjk7@mit.edu

ABSTRACT: This proposal sets up a challenge for 6th Learning Analytics Hackaton @ LAK 2020 to design and implement a Learning Analytics algorithm that can measure the spatial reasoning abilities of the player while interacting with the geometry game Shadowspect.

Keywords: Game-based Assessment, Learning Games, Learning Analytics, Spatial Reasoning

1 BACKGROUND

Plenty of studies have shown the potential of using games for learning and also as assessment tools. Game-based Assessment has emerged as an alternative and less intrusive method than traditional assessment, in which we can indirectly measure students' abilities without interrupting the learning experience. In this challenge we use Shadowspect, a geometry game explicitly build for assessment purposes. Shadowspect sessions consist of a series of puzzles, where each one is composed of three orthogonal views of a figure, where each figure is composed of a series of 3D geometric primitives. Participants build a 3D figure by using the 3D game environment prototype to solve the puzzles, or to create imaginative structures in the game's sandbox mode ([see a video online](#)). This challenge focuses on measuring the complex cognitive skill of spatial reasoning, based on the data generated through the interaction of students with Shadowspect puzzles.

2 RESEARCH QUESTION

The challenge proposes to algorithmically measure spatial reasoning, and thus the objective is:

- Design and implementation of a data-driven algorithm that can measure the spatial reasoning abilities of students based on their previous interaction with Shadowspect.

3 EXPECTED OUTCOMES

The ideal outcome of this challenge would be an algorithm applicable to Shadowspect data that could provide evidence of the spatial reasoning capabilities of a student, and that this joint work could be the seed of a publication about this topic.

Smartwatches in Education

Daniele Di Mitri

Open University of The Netherlands
daniele.dimitri@ou.nl

Khaleel Asyraaf Mat Sanusi

Open University of The Netherlands
khaleel.asyraaf@ou.nl

Jan Schneider

DIPF - Leibniz Institute for Research and Information in Education
schneider.jan@dipf.de

ABSTRACT: the use of connected wearable devices is increasing rapidly and it is forecasted to be over one billion in 2022. While wearable smartwatches becoming increasingly popular, they introduce new technological affordances. In this LAKathon challenge, we want to investigate the use of smartwatches to support practical tasks both in educational and professional environments.

Keywords: multimodal learning analytics, wearables, smartwatches, sensor-based learning

1 BACKGROUND

The permeation of connected wearable devices is increasing rapidly, and it is forecasted to be over one billion in 2022. Wearable devices are electronics that can be worn on the body such as smartwatches, wristbands or earbuds. Typical commercial smartwatches embed powerful microchips and sensors and can connect via the cellular network, Wi-Fi, NFC or Bluetooth. From being primarily chosen by athletes and fitness enthusiasts, smartwatches are being progressively adopted also by the general population. Smartwatches allow collecting physiological data, such as step counting, heart-rate tracking and sleep monitoring. They also provide a hands-free interface, which enables the user to stream music, receive notifications, interact with conversational agents having their hands free. Compared to smartphones, smartwatches have smaller screen size and a limited Graphic User Interface, which makes the smartwatch less “task-dominant” in day-to-day tasks and thus less alienating as compared to smartphones. This makes the smartwatch a better-suited device for supporting practical tasks such as doing fitness or cooking.

2 RESEARCH QUESTIONS

Building on the previous research conducted in the field Multimodal Learning Analytics, leveraging the Learning Hub (Schneider, Di Mitri, Limbu, & Drachsler, 2018) for multimodal data collection and the Visual Inspection Tool for data annotation (Di Mitri, Schneider, Klemke, Specht, & Drachsler, 2019) in this LAKathon challenge, we would like to explore can new technological affordances introduced by the smartwatches can be leveraged in education.

3 EXPECTED OUTCOME

A conceptual design of a smartwatch application which continuously can 1) collect sensor data, 2) ask user reports and 3) return valuable information to the user for optimising a particular task.

REFERENCES

- Di Mitri, D., Schneider, J., Klemke, R., Specht, M., & Drachsler, H. (2019). Read Between the Lines: An Annotation Tool for Multimodal Data for Learning. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge - LAK19* (pp. 51–60). New York, NY, USA
- Schneider, J., Di Mitri, D., Limbu, B., & Drachsler, H. (2018). Multimodal Learning Hub: A Tool for Capturing Customizable Multimodal Learning Experiences. In *Lecture Notes in Computer Science* (Vol. 11082 LNCS, pp. 45–58). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-98572-5_4

Adoption, Adaptation and Pilots of Learning Analytics for Latin American Higher Education Institutions

Pedro J. Muñoz-Merino, Carlos Delgado Kloos

Universidad Carlos III de Madrid, Spain
pedmume@it.uc3m.es, cdk@it.uc3m.es

Yi-Shan Tsai

University of Edinburgh, United Kingdom
Yi-Shan.Tsai@ed.ac.uk

Dragan Gasevic

Monash University, Australia
dragan.gasevic@monash.edu

Katrien Verbert

Katholieke Universiteit Leuven, Belgium
katrien.verbert@cs.kuleuven.be

Mar Pérez-Sanagustín*, Isabel Hilliger

*Université Paul Sabatier Toulouse III (IRIT), Pontificia Universidad Católica de Chile, Chile
mar.perez-sanagustin@irit.fr, ihillige@ing.puc.cl

Miguel Ángel Zúñiga Prieto

Universidad de Cuenca, Ecuador
miguel.zunigap@ucuenca.edu.ec

Margarita Ortiz

Escuela Superior Politécnica del Litoral, Ecuador
margarita.ortiz@cti.espol.edu.ec

Eliana Scheihing

Universidad Austral de Chile, Chile
escheihi@inf.uach.cl

ABSTRACT: Learning analytics needs to make more progress in specific world regions such as Latin America. This workshop proposal aims at presenting different works on the adoption, use, adaptation and evaluation of learning analytics in the Latin American region. The workshop will present results from the LALA European funded project in addition to works of other participants invited through an open call for papers. The main objectives are to better understand the level of development of learning analytics in Latin America, present a framework for the adoption of learning analytics in Latin America, present how different learning analytics tools have been adapted to the Latin American region, and to learn about experiences and evaluations in this region. With this WS, we also aim at reinforcing the community of learning analytics in Latin America and connect this community with others in different regions of the world.

Keywords: Higher education, learning analytics, Latin America, adoption, tools, pilots

1 BACKGROUND

Learning analytics research has made a notable progress in the last decade. However, from a regional perspective, not all the regions are using and/or researching learning analytics at the same level. Regions such as North America, Europe or Australia have developed learning analytics in a considerable way. However, in regions such as Latin America there is a need to build local capacity to use learning analytics. In this context, the application of learning analytics in Latin America represents a great opportunity (Ochoa, 2019). At present, the interest of the application of learning analytics in Latin America is increasing. Proof of that growing interest is e.g. the current special issue on learning analytics adoption in Latin America of British Journal of Educational Technology (Pontual Falcão, Ferreira, Rodrigues, 2019)

From an institutional point of view, there is a need to analyse how learning analytics can be adopted in higher education institutions (HEIs) taking into account the contextual particularities of Latin American HEIs regarding culture, privacy, ethics or technical infrastructure. In addition, there is a need to adapt different learning analytics tools and researches to meet the needs in the local contexts, rather than treating learning analytics as generalizable solutions.

The LALA project (LALA project, 2017) is an Erasmus + project which seeks to build local capacity in Latin American HEIs to design and implement learning analytic tools in order to improve their learning processes. The main outcomes and results of the Project are: 1) developing a framework that describes the methodological, technical, institutional, ethical and communal aspects required for the adoption of learning analytics for Latin American institutions; 2) adapting two learning analytics tools (a counselling tool for visual analytics and an early dropout prediction tool) for Latin American institutions; 3) piloting the two learning analytics tools in different Latin American institutions; and 4) creating a community of learning analytics in Latin America.

The LALA proposes the LALA framework (Pérez-Sanagustín et al., 2018) to help Latin American HEIs to adopt learning analytics. This framework builds on the SHEILA framework (Tsai et al., 2018) proposed in the SHEILA Erasmus + European project.

So far, the LALA project team has already organized similar events in several Latin American countries. For example, the 1st LALA conference was organized in Guayaquil (Ecuador) and the proceedings of the event are available (Ochoa & Pérez-Sanagustín, 2018), and the 2nd LALA conference was organized in Valdivia (Chile) and the proceedings of the event are available (Scheihing, Guerra, Henríquez, Oliares, & Muñoz-Merino, 2019).

2 OBJECTIVES

The objectives of this workshop are as the following:

- To learn about the different initiatives about learning analytics that are currently being conducted in Latin America.

- To introduce the results and outcomes of the LALA European Union funded project about the adoption and adaptation of learning analytics in Latin America as well as piloting.
- Spread and discuss the knowledge about applications and experiences of learning analytics in Latin America.
- Reinforce the community of Learning Analytics in Latin America.
- Promote networking between the learning analytics community in Latin America and other learning analytics communities around the world.

3 GENERAL ORGANIZATIONAL DETAILS

The expectations of this workshop are described below:

- Type of event: mini-track / symposium. The workshop includes a series of presentations about the outputs of the LALA project, including the framework for the adoption of learning analytics in Latin America, the adaptation of two different learning analytics tools for different Latin America partners and pilot experiences. In addition, there is an open call to present works of learning analytics adoption in Latin America with a review process conducted by a committee, including the workshop organizers and other experts.
- Type of participation: 'open' workshop (i.e., any interested delegate may register to attend).
- Expected number of participants: 25. It is expected that 10 participants will come from the Project team, and 15 will be external participants.
- Planned dissemination activities to recruit attendants: the workshop will be disseminated using the Project webpage, Twitter and Facebook accounts. In addition, we will distribute the workshop information through the community mailing list, which is subscribed by more than 150 institutional members.
- Required equipment: projector.

4 PROPOSED SCHEDULE

This is a Full-day workshop. The workshop is planned to be organized into 3 + 3 hours. In this workshop, we combine presentations of the LALA project with presentations from open submissions. The proposed schedule includes:

- Presentation of the LALA project and achieved general objectives, including the LALA framework, adaptation of dashboards for counselling tools, adaptation of early dropout prediction systems and presentations of pilots (2 hours 30 mins.)
- Workshop paper presentations from the open call with a review process (2 hours 30 mins.).
- Discussion (1 hour)

5 PROPOSED CALL FOR PAPERS

An open Call for Papers was done for this workshop. Each paper proposal has been reviewed by at least two experts. Proceedings of accepted workshop papers are expected to be published in CEUR (<http://ceur-ws.org/>). Papers that address any aspect of learning analytics in Latin America are welcomed, including among others:

- Reviews of the state of the art of learning analytics in Latin America.
- Adoption of learning analytics in Higher Education Institutions in Latin America.
- Adaptation of learning analytics tools for Latin American institutions.
- Experiences and evaluation of learning analytics in Latin America.

ACKNOWLEDGMENTS

Work funded by the LALA project (grant no. 586120-EPP-1-2017-1-ES-EPPKA2-CBHE-JP). This project has been funded with support from the European Commission. This communication reflects the views only of the authors, and the Commission and the Agency cannot be held responsible for any use which may be made of the information contained therein.

REFERENCES

- LALA project (2017), <https://lalaproject.org/>
- Ochoa, X. (2019). Learning analytics in Latin America present an opportunity not to be missed. *Nature human behaviour*, 3(1), 6.
- Ochoa, X., Pérez-Sanagustín, M.P. (2018), Proceedings of the 1st Latin American Workshop on Learning Analytics (LALA 2018), Guayaquil, Ecuador, July 9th, 2018, <http://ceur-ws.org/Vol-2231/>
- Pérez-Sanagustín, M., Hilliger, I., Maldonado, J., Pérez, R., Ramírez, L., Muñoz-Merino, P. J., Tsai, Y., Ortiz, M., Broos, T., Zuñiga-Prieto, M.A., Sheihing, E., WhitelockWainright, A. (2018), LALA Framework, https://www.lalaproject.org/wp-content/uploads/2019/04/LALA_framework_Spanish.pdf
- Pontual Falcão, T., Ferreira, R., Rodrigues, R.L. (2019), Applications of Learning Analytics in Latin America, *British Journal of Educational Technology*, https://onlinelibrary.wiley.com/page/journal/14678535/homepage/bjet_special_issues.htm
- Scheihing, E., Guerra, J., Henríquez, V., Olivares, C., Muñoz-Merino, P.J. (2019) Proceedings of the 2nd Latin American Workshop on Learning Analytics (LALA 2019), Valdivia, Chile, March 18-19, 2019, <http://ceur-ws.org/Vol-2425/>
- Tsai, Yi-shan, Gasevic, D., Whitelock-Wainwright, A., Muñoz-Merino, P. J., Moreno-Marcos, P. M., Rubio Fernández, A., ... Kollom, K. (2018). SHEILA Supporting Higher Education to Integrate Learning Analytics.

LAK2020 2nd Workshop on Learning Analytic and Services to Support Personalized Learning and Assessment at Scale

Alina von Davier, Michael Yudelson

ACTNext by ACT, Inc

{alina.vondavier, michael.yudelson}@act.org

John Stamper

Carnegie Mellon University

kodinger@cmu.edu

Steve Ritter

Carnegie Learning, Inc

sritter@carnegielearning.com

Peter Brusilovsky

University of Pittsburgh

peterb@pitt.edu

ABSTRACT: The workshop will focus on the conceptual frameworks, algorithms, and analytical approaches that fuel modern learning and assessment systems (LASs). Modern systems of computer-supported education have matured often in separate siloed fields of research and there is a need for forming an overarching conceptual schema covering a wide array of approaches implemented in the existing and future products supporting the educational process at various stages: from the operational recommendation and adaptation to the offline investigation. Compartmentalization of relevant fields of research is limiting further betterment of the LASs and this workshop is focusing on overcoming a what now seems as an artificial separation.

Keywords: approaches to adaptation, personalization, and recommendation; learning analytics as [micro] service; standardization of data exchange; validity of assessment; product improvement; architecture and scaling.

1 BACKGROUND

The focus of this workshop is on multidisciplinary research in the areas of personalization and adaptation of digital education and assessment tools. Recent developments and the prior instance of this workshop indicate interest in rethinking learning and assessment systems that have largely been developing separately and seldom thought as complimenting parts of a unified Learning Assessment System (LAS). Such compartmentalization resulted in incremental improvement of the systems we have. The educators often rely on formative assessments (e.g., weekly quizzes) that operationally reflect upon overall classroom standing with respect of conceptual mastery and students' relative progress. There is also a growing interest in performance assessments and learning that are individualized and adaptive and are carried out in a standardized ubiquitous manner. Traditional approaches are unable to fully explain why students perform as they do and are not yet suited to

measure increasingly important constructs like behavior, affect, or collaboration. The desire of the field at large is to make progress towards LASs that are educationally effective, reflect realistic educational goals, accommodate student collaboration, and provide reliable instructional support for teachers. The early attempts to create such systems demonstrate a great potential. However, these LASs come with many challenges in terms of measurement, operation, and unification and standardization. Recent advances in applied machine learning (ML) offer opportunities to address these challenges by aggregating and analyzing the Big Data that is accumulated when students interact with LASs. Among other approaches are those that structure data into various forms of learner record stores, align instructional content and the assessment content across theoretically or empirically defined knowledge component schema and standards such as the Common Core State Standards or the New Generation State Standards. Analytical platforms and service providers that offer operational and post-hoc investigation support for adaptation of learning paths and assessment delivery.

This workshop would be in its second iteration following a successful execution at LAK 2019 in Arizona where it was sold out. It provides a venue to researchers that have been working on multiple components of LASs for an extended period of time to share their experience and their findings in the area of framework building and scaling the personalization and adaptation approaches. This time, we would like to have an open submission process and, while encouraging a group of teams to submit, let a wider population of researchers to contribute. We are reserving an option to have several invited keynotes as well. We will focus on finding the common ground between the approaches presented and would work toward advancing the agenda of LASs further.

2 ORGANIZATION

The workshop will be in the form of Symposium and is intended as full-day with an estimated number of 8-12 oral presentations (openly submitted and invited keynotes). The organizers will moderate the course of the workshop as well as emerging discussions. Organizers will provide an introduction to the workshop, will lead the discussion after each oral presentation, and summarize the workshop. Workshop attendees will be actively involved in post-presentation Q&A and other discussions throughout the day. The workshop will follow an open attendance model: any conference delegate may register to attend. Each speaker will present their ongoing research and will give a brief overview of the state-of-the-art methods and applications in their respective field.

A tentative schedule is to consist of 4 sessions, 3 breaks for coffee and lunch and a closing discussion (see Table 1).

Table 1: A tentative workshop schedule.

Activity	Time
Registration	8:30-5:00
Introductory remarks	8:50-9:00
Session 1 (3 presentations)	9:00-10:30
Presentation 1	9:00-9:30
Presentation 2	9:30-10:00
Presentation 3	10:00-10:30
Morning Tea/Coffee	10:30-11:00
Session 2 (2 presentations)	11:00-12:00

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

Activity	Time
Presentation 4	11:00-11:30
Presentation 5	11:30-12:00
Presentation 6	12:00-12:30
Lunch	12:30-1:30
Session 3 (5 presentations or posters)	1:30-3:30
Presentation 7	1:30-2:00
Presentation 8	2:00-2:30
Presentation 9	2:30-3:00
Presentation 10	3:00-3:30
Afternoon Tea/Coffee	3:30-3:50
Session 4 (2 presentations)	3:50-4:50
Presentation 11	3:50-4:30
Presentation 12	4:30-4:50
Closing Remarks	4:50-5:00

3 PAPER SUBMISSION AND REVIEW

ACTNext will create and maintain the workshop website that will sustain well after workshop is over. Submissions will be accepted online. Questions should be directed to Michael Yudelson (michael.yudelson@act.org) with the subject set to “LAK 2020 Adaptive Engines Workshop”. All submissions will be judged on their novelty, conceptual and opinionated reflection on the topic of the workshop, and potential impact on the adaptive and personalized delivery of educational content and assessment.

Submission deadline: December 20, 2019

Notifications to authors: January 5, 2020

Author’s instructions: https://lak20.solaresearch.org/wp-content/uploads/2019/07/LAK20_Companion_Proceedings_Template.doc

Workshop website: <http://actnext.info/LASSPLAS20/index.html>

4 INTENDED OUTCOMES

We believe multidisciplinary research and collaboration is key to developing the next generation learning and assessment systems that amass a critical set of adaptive support methods to cater to student needs and bring the sophistication and pedagogical nuances of a good teacher. This workshop would provide a forum for the sharing of knowledge and ideas across disciplines including computational psychometrics, adaptive learning and testing, and learning analytics, machine learning, educational measurement, and natural language processing. The research is relevant and timely for advances in learning and performance assessment simulation systems and collaborative LASs. We expect that by bringing together some of the best minds in these fields, we will be able to further the state of the art and generate an increasing interest and excitement in this area.

Teacher-sourcing semantic information in a Physics blended-learning environment

Giora Alexandron

Weizmann Institute of Science
Giora.alexandron@weizmann.ac.il

Elad Yacobson

Weizmann Institute of Science
Elad.yacobson@weizmann.ac.il

Asaf Bar Yosef

Weizmann Institute of Science
Asafbj@weizmann.ac.il

Eliran Hen

Weizmann Institute of Science
Heneliran@gmail.com

ABSTRACT: Personalized learning environments rely on repositories of digital learning materials, and on meta-data that provides semantic information about the digital content. Semantic information is typically generated by domain experts, but this process is very time consuming, and it fails to address the dynamic nature of the semantic information, the content, and the contexts in which it is used. In addition, experts may fail to capture semantic properties that are not within their area of expertise. Overall, expert-based semantic generation processes do not scale, and produce limited information. Thus, the goal of our research is to study means to scale the process of collecting and updating semantic information, using teacher- and learner-sourcing. As a proof-of-concept, we conducted a pilot experiment with two groups of physics teachers who are using an Open Educational Repository. The main goal was assessing the *quality* of the semantic information that the teacher-sourcing produces, and factors affecting it. Results showed that teachers can tag items relatively accurately (Cohen's kappa: 0.56) even without having a full knowledge of the taxonomy from which the tags are taken. In addition, verbal analysis of teachers' discussions yielded interesting insights about the different factors that foster effective teacher-sourcing, and its potential contribution to teachers' professional development.

Keywords: Personalized Learning, Semantic Information, Teacher Sourcing

1 INTRODUCTION

Personalized learning environments rely on repositories of digital learning materials (e.g., interactive questions, online labs, videos), and on meta-data that provides rich semantic information about the digital content. The semantic information is fundamental to the ability of AI agents to make 'intelligent' decisions such as recommending content to learners, to assist teachers in searching & discovering of learning resources, and for re-using and sharing materials between contexts (Aroyo & Dicheva, 2004; Bittencourt et al., 2012). While high-quality digital content is in many cases readily available on the web, it is the semantic information that is usually missing, inadequate, or partial. Thus, having scalable processes for generating high-quality semantic information can contribute significantly to the development of personalized learning environments. Semantic information is typically generated by domain experts, but this process is very time consuming, and the experts may fail to capture semantic properties that are not within their area of expertise (McCalla, 2004). In addition, the content repository and the context in which it is used are dynamic, requiring frequent revisions and updates. Overall, expert-based semantic generation processes do not scale, and produce limited information.

The high-level goal of our research is to study crowdsourcing (more accurately: teacher- and learner sourcing, which are the terms that we use hereafter) as means to scale the process of collecting and updating semantic information.

2 RESEARCH QUESTIONS

The specific study described here aims to address the following research questions (RQs):

RQ1: How do the characteristics of the teacher-sourcing task affect teachers' ability to provide accurate semantic information? Specifically – can teachers tag questions without having a full knowledge of the taxonomy from which the tags are taken? To which resolution do teachers need to go in analyzing the questions in order to provide accurate tags?

RQ2: What is the level of agreement between expert and teacher-sourced tagging?

RQ3: Does participating in teacher-sourcing contribute to teachers' professional development, and specifically, to their ability to use semantic information in order to personalize their instruction?

The rationale behind these questions is understanding the factors that affect the quality of the teacher-sourcing outcome, in order to find a *modus operandi* that optimizes the balance between teachers' time and effort and the accuracy of the results. For example, with respect to RQ1, teachers might not be well familiar with the details of the taxonomy, and requiring them to master it may demand time and effort that will significantly reduce participation. Thus, we study questioning modes that can be done with partial knowledge of the taxonomy.

3 LEARNING ENVIRONMENT

3.1 *PeTeL*

PeTeL is both a shared repository of open educational resources (OER), and an LMS that also includes social network features and learning analytics tools. It is developed within the Department of Science Teaching at Weizmann Institute of Science, with the goal of providing STEM teachers with a blended learning environment for personalized instruction. PeTeL is divided into separate modules for each subject matter: Biology, Chemistry and Physics. The Physics module is currently being used by approximately 200 teachers and 7000 high school students.

3.2 *Semantic Tagging and Taxonomy*

Each interactive activity in the Physics OER in PeTeL is tagged with semantic information in order to support search & select. The tags capture two types of semantics: content-knowledge and general skills. The taxonomy of the content-knowledge tags is based on the Physics curriculum as determined by the Ministry of Education (MoE). Each tag describes a fine-grained concept, and it is assigned to questions that require this concept. For example, content-knowledge tags can contain the following information: "Magnetic field lines exit the north pole from the outside and enter in the south pole", "The magnetic force direction is perpendicular to the direction of the magnetic field and the direction of the electric current", etc. The skill tags describe the general skills required to solve the question, and are orthogonal to the content-knowledge. For example, skill tags can contain the following information: "3D perception", "translating from graphic to algebraic representation", "extracting information from text", "interpreting diagrams and graphs", etc.

In the experiments described in this paper we used only the content-knowledge tags.

3.3 Refining the MoE taxonomy: Magnetism as a test case

To support detailed analysis of student learning needs, and tailoring personalized instruction that addresses them, it was evident from the beginning that a much richer and fine-grained semantic information is required on the items. The Physics R&D team decided to start with the topic of Magnetism as a test-case. This topic captures about 10% of the curriculum, and is pretty much independent in terms of prerequisites.

The MoE taxonomy for Magnetism has two levels. The first is divided into six subsections: (1) magnetic fields of magnets and electric currents; (2) the influence of the magnetic field on an electric current; (3) the connection between the magnetic field and its sources; (4) the force between parallel currents and the definition of Amper; (5) the force working on an electric charge moving through a magnetic field; and (6) implementations of the magnetic force. Each of these six subsections is then divided into 2 – 6 sub(sub)sections.

Under these two levels, a team of domain experts (teachers and researchers from the Physics Education Research Group) added another level, yielding a 3-level taxonomy. Once the fine-grained taxonomy was completed and reviewed, the ‘tagging’ team tagged 250 questions in Magnetism with this taxonomy. Interpreting the leaves of the fine-grained taxonomy as concepts or skills, this yields a Q-matrix that maps each question to the skills that are required for solving it (Tatsuoka, 1983).

Overall, the process of developing the detailed taxonomy, and mapping the items into it, took considerable human effort (a few weeks of work, spread over several months).

4 METHODOLOGY

4.1 Experimental Setup.

We conducted two pilot experiments with two groups of Physics teachers – eight teachers in the first experiment and seven teachers in the second experiment. The first experiment was conducted face-to-face, whereas the second experiment was held online. All of the fifteen teachers who took part in these experiments use PeTeL in their classrooms. Within these experiments, the teachers were requested to tag questions from Magnetism into the detailed taxonomy.

4.2 First Experiment

In the first experiment, eight teachers were presented each with three questions from PeTeL. Each question contains a picture or diagram of a certain Physics situation (e.g. a particle moving through a magnetic field, or an electric circuit), and a question regarding that diagram (See example in Figure 1). For each question i , the teachers were presented with four tags from the Magnetism taxonomy. Among these, two were correct tags of i (as determined by the domain expert team), and two were incorrect ones. Then, for each tag, the teachers were requested to decide whether it applies to i .

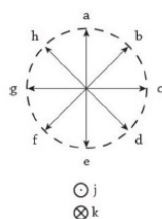
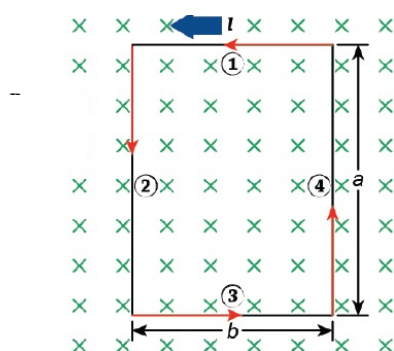
After the teachers completed going through the three questions, they divided into pairs and compared their decisions. If a disagreement was found concerning the necessity of a specific content-knowledge

tag for a certain question, the teachers discussed it, trying to reach an agreement. Finally, a group discussion was held, in which the teachers shared insights and perspectives.

4.3 Second Experiment

The second experiment took place about a month after the first one, with a different group of seven Physics teachers. It followed the same protocol, except for three differences – first, it was held online and not face-to-face; Second, the teachers were presented with the same two questions (vs. three, with partial overlap, in the first experiment); Third, the group discussion was held immediately after each teacher worked individually, without the teachers working in pairs.

A rectangular frame with sides a and b , is shown in the following diagram. An electric current is flowing through each of the frame's sides in counter-clockwise direction. The frame is located in a magnetic field entering the page's plain, as shown in the diagram.



- * what is the direction of the magnetic force working on side a of the frame?
- * what is the direction of the magnetic force working on side b of the frame?
- * what is the direction of the magnetic force working on side c of the frame?
- * what is the direction of the magnetic force working on side d of the frame?

Does solving this question require the following concepts?

- * The magnetic field creates a force over a current-carrying wire - **yes / no**
- * Effect of different parameters on the force: magnitude of magnetic field and of current - **yes / no**
- * The magnetic force's direction is vertical to the direction of the magnetic field and to the direction of electric charges - **yes / no**
- * Effect of different parameters on the force: the angle between the direction of the field and the direction of the current - **yes / no**

Figure 1: Example of PeTeL question with 4 content-knowledge tags (translated from Hebrew)

4.4 Data analysis

Data include the teachers' answers about the relevance of each tag, recording of the pair and group discussions, and expert tagging as ground truth. We used a mixed-method approach: Teachers' answers were analyzed quantitatively, and their discussions were analyzed qualitatively.

5 RESULTS

5.1 First Experiment

As described before, data were collected from eight teachers, each presented with three questions and four candidate tags per question. Overall, we received 95 responses (one response was missing).

The results are presented in Table 1:

Question	Tag	Domain Expert	Teacher1	Teacher 2	Teacher 3	Teacher 4	Teacher 5	Teacher 6	Teacher 7	Teacher 8
Q1	A	Yes	N	N						
	B	Yes	Y	Y						
	C	No	N	N						
	D	No	N	N						
Q2	A	Yes	N	N	Y	Y				
	B	No	N	N	N	N				
	C	Yes	Y	Y	Y	Y				
	D	No	N	N	Y	N				
Q3	A	No	N	N					Y	Y
	B	Yes	Y	Y					Y	Y
	C	Yes	Y	Y					Y	Y
	D	No	N	N					N	N
Q4	A	No			N	N				
	B	Yes			N					
	C	No			N	N				
	D	Yes			Y	Y				
Q5	A	No			N	N				
	B	Yes			Y	Y				
	C	Yes			N	N				
	D	No			Y	Y				
Q6	A	Yes					Y	Y	Y	Y
	B	No					N	N	N	N
	C	Yes					Y	Y	Y	Y
	D	No					N	N	N	N
Q7	A	Yes					Y	Y		
	B	No					N	N		
	C	No					N	N		
	D	Yes					Y	Y		
Q8	A	Yes					N	N	N	N
	B	No					N	N	Y	Y
	C	No					N	N	N	Y
	D	Yes					N	N	Y	Y

Table 1: Teachers' Responses in the first experiment

In 74 out of 95 responses, the teachers agreed with the domain expert as to whether the content-knowledge described in the tag is required for solving the question (**78% agreement, Cohen's kappa: 0.56**).

Confusion matrix:

		Domain expert tagging	
		Yes	No
Teachers' tagging	Yes	34	8
	No	13	40

No substantial differences were found between the 8 teachers in their level of agreement with the domain expert.

5.2 Second Experiment

In the second experiment, data were collected from seven teachers, each presented with two questions and four tags per question. A total of 56 responses were collected.

The results are presented in Table 2:

Question	Tag	Domain Expert	Teacher1	Teacher 2	Teacher 3	Teacher 4	Teacher 5	Teacher 6	Teacher 7
Q1	A	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	B	No	No	Yes	Yes	Yes	No	Yes	Yes
	C	No	Yes	Yes	No	No	Yes	No	Yes
	D	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Q2	A	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	B	No	Yes	Yes	No	Yes	No	No	No
	C	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	D	No	No	No	No	No	No	Yes	No

Table 2: Teachers' Responses in the second experiment

In 43 out of 56 responses, the teachers agreed with the domain expert as to whether the content-knowledge described in the tag was required for solving the questions (**77% agreement, Cohen's kappa: 0.54**).

Confusion matrix:

		Domain expert tagging	
		Yes	No
Teachers' tagging	Yes	28	13
	No	0	15

No substantial differences were found between the 7 teachers in their level of agreement with the domain expert.

5.3 Teachers' Discussions

Three main categories emerged from the verbal analysis: (a) what teachers need in order to perform the tagging task; (b) the manner in which the tagging was done; and (c) the potential contribution of participating in tagging on the teachers' professional development.

5.3.1 What teachers needed for performing the tagging assignment:

With regard to RQ1, during their individual work, in the pair discussions, and in the group discussions, no questions or dilemmas were raised concerning the taxonomy from which the tags were taken. We believe that this results from the presentation of the questions about the required content-knowledge tags as YES/NO questions. In addition, in most cases, the teachers solved the questions and then analyzed their solutions before deciding which tags are relevant to each question. Roughly, this took several minutes per question.

Another issue that was noticed in the second experiment was that teachers' **motivation** affected the quality of their tagging. The second experiment started at 21:40 PM, after over an hour of other activities and discussions concerning different features of PeTeL. By that time, the teachers were tired

and seemed less cooperative. After they performed the tagging assignment, one of the teachers was asked about one of his responses. He marked a certain tag as required for the solution of the question, while all the other teachers agreed that it was not. When asked about it, he replied: "I was tired, I only looked at the question and the tag superficially, and got confused. Of-course I understand now that this tag is not relevant to the solution of this question".

5.3.2 *The manner in which the tagging was done*

Following their tagging assignment in the first experiment, the teachers emphasized two important aspects of the tagging system and the task definition: First, that it should be made clear that the content-knowledge described in the tag is *required for solving the question*. For example, one question dealt with a particle entering a magnetic field. The question was "What would be the predicted course of movement of the particle through the magnetic field?". Four possible answers (choices) were presented, all depicting some form of a circular movement. Then the teachers were asked whether the tag "A charge particle entering a magnetic field vertically will perform a circular movement" is required for solving the question. Most teachers agreed that although this particular content-knowledge is **relevant** to this question, it is **not required** for solving it since all of the four choices include circular movement (i.e., this information is given in the question).

Another issue raised by the teachers concerned the *scale* of relevance: The teachers mentioned that concepts may be primary to the question at hand, or secondary, namely, that there is a *scale of relevance*. In addition, some of the concepts are mandatory, while others are relevant, but are not mandatory (e.g., if there are several correct solutions).

5.3.3 *potential contribution to professional development*

Tagging questions with the content-knowledge that is required for their solution is a reflective activity by nature. The teachers must first solve the question, and then actively ask themselves what kind of knowledge they used during the process. Since reflection is considered to be a tool for advancing learning and understanding, we wanted to see whether the teachers felt that the tagging assignment they were asked to accomplish during the experiments contributed to their professional development (PD). Obviously, PD is a longitudinal process, and its outcomes should be measured in an objective manner (e.g., the effect of tagging on teachers' awareness of individual needs among students, and their ability to model and treat those differences in an effective manner). At this stage the results are only anecdotal and subjective, but encouraging, as concluded spontaneously by one of the teachers who participated in the first experiment: "It is very useful for a teacher to do what we just did. A teacher who is involved in tagging will become an excellent teacher... It is great for us as teachers".

6 CONCLUSIONS AND FUTURE RESEARCH

This paper describes two small-scale experiments with Physics teachers, in which we piloted several teacher-sourcing task definitions, and their influence on the quality of the teacher-sourcing process. The results of these experiments suggest that when the tagging task is formulated in a certain way (e.g., "yes/no" questions), teachers can tag items relatively accurately (Cohen's kappa: 0.56) without being trained on the taxonomy from which the tags are taken. In addition, the results show that teachers tend to solve the questions before tagging them. Since this requires effort that may eventually decrease motivation to participate, one of the lessons learned is that it might be better to present teacher-sourcing tasks *after* the teacher ran the learning activity in her class (as opposed to

when selecting it). Regarding motivation, the results also suggest that the *quality* of teacher-sourcing can be significantly affected by teachers' motivation to participate in the tagging activity. One of the design lessons learned from this is that positive incentives, rather than negative ones (e.g., requiring participation for receiving access to materials) are more likely to produce quality results.

On the next step, we intend to run this experiment on a larger scale, using a technological tool to teacher-source semantic information from a much larger pool of teachers.

REFERENCES

- Aroyo, L., & Dicheva, D. (2004). The New Challenges for E-learning: The Educational Semantic Web. *Educational Technology & Society*, 7 (4), 59-69.
- Bittencourt, I., Isotani, S., Costa, E., Mizoguchi, R., and Aprigio Veloso, R. (2012). *Research Directions on Semantic Web and Education*.
- McCalla, G. (2004). The ecological approach to the design of e-learning environments: Purpose-based capture and use of information about learners. *Journal of Interactive Media in Education*.
- Tatsuoka K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*.

Understanding Reflective Writing Criteria in Computer Science Education from CS Educators in Higher Education

Huda Alrashidi^{1,4}, Thomas Daniel Ullmann², Samiah Ghounaim³, Mike Joy¹

Computer Science Department, University of Warwick¹

Institution Institute of Educational Technology, The Open University²

Centre for Applied Linguistics, University of Warwick³

h.alrashidi@warwick.ac.uk⁴

ABSTRACT: Reflective Writing has many benefits to increase students' awareness of the ways in which they are gaining insight into their learning processes. However, there is a lack of studies that deal with reflective writing analysis frameworks in the context of computer science (CS) education. The overall goal of this present research is to develop a Learning Analytics (LA) tool which can automatically detect the categories of a reflective writing framework (RWF) present in a text to assess the student authors' reflective writing in relation to CS. Here, we present the RWF that we developed based on an expert questionnaire. Findings from the responses to the open-ended questions identified (a) three reflection levels, and (b) seven indicators relating to these and to reflective writing generally –in CS.

Keywords: Reflective Writing, Computer Science, Reflection, Reflection Detection, Reflective Writing Analytics, Learning Analytics.

1 INTRODUCTION

Learning Analytics (LA) is increasingly gaining attention in relation to educational technology. For example, there are LA tools that aim to support reflection by analyzing (Ullmann, 2019) and providing feedback (Gibson et al., 2017) with respect to Reflective texts. Reflective writing is an important skill as it offers critical thinking and enhances awareness of the learning processes required in higher education. In spite of several studies carried out in reflective writing based on medical and education fields, there is a scarcity of an exploratory study to integrate reflective writing in CS education. This, in turn, would depict the aim of the current practice to identify the criterion of using reflection within CS education in order to support the implementation of automated reflective writing analysis.

George (2002) claimed, *“reflection in scientific disciplines may be different in type to the type of reflections made in humanities because of the nature of the underlying knowledge”*. She also mentioned that the underlying knowledge is declarative in humanities and social sciences, which is composed of facts while problem-solving and reasoning are not necessary to add during the event or situation. In CS education, reflection is used to improve students' awareness in order to learn from a situation such as how to deal with a sequence of steps to reach a certain goal and how to identify the roots of problems rather than their feelings during that situation (Chng, 2018).

In terms of CS education, as stated by Fekete, Kay, Kingston, and Wimalaratne (2000), *“ reflection is worth encouraging, for its indirect effect on the technical skills and knowledge which are our ultimate purpose in teaching Computer Science”*. Technical skills are, of course, at the core of CS education, and it consists of *“ thought processes involved in formulating a problem and expressing its solution(s) in such a way that a computer—human or machine—can effectively carry out”* (Wing, 2014). That technical skill has been reported to have the following components: problem formulation and

understanding, literature analysis, work planning. More importantly, the technical skills also required to produce implementations and report on the results accurately. These components are based on both cognitive and metacognitive abilities.

It is essential to have clear measures for assessing reflective writing which is based on the reflective process as expressed by the written text. When reflective writing is analyzed manually, this makes it a challenging and time-consuming task that involves content analysis of students' texts. The main issue carried out in this research is the atomization of reflective writing analysis in CS education to overcome the difficulties in manual process. After conducting a project related to teaching and assessment of reflective writing, (Ryan, 2011) indicated that *"Many academics lack the meta-language to identify or explain what they regard as key elements of deep reflective writing"*. They are therefore unable either to give clear directions to students about how to approach a reflective writing task or to justify the marks that they give to students' assignments".

Developing the process of assessing reflection in writing is not fully covered (Poldner, Simons, Wijngaards, & Van der Schaaf, 2012) and suffers from a lack of dedicated researchers on reflection assessment (Ryan, 2011; Shum et al., 2016). The insufficient work on the automatic analysis of reflective writing in text for education (Corich, 2011; Liu, Shum, Mantzourani, & Lucas, 2019; Moseley et al., 2004) led to producing more work that focus on such areas. Undetermined problems can be solved using reflective thinking (Thorpe, 2004). The analysis of reflective writing is necessary for educational practice for educators as it enables the assessment of the writer.

An automatic reflective writing analysis has started recently, which affirmed that different scientific fields required different attention. Thus, the automated reflective writing analysis at various levels, such as higher education was addressed in undergraduate studies (Gibson et al., 2017; Kovanović et al., 2018; Shum et al., 2016). Similarly, reflective writing in various scientific fields can gain many benefits compared to those of a general form.

The wider goal of this research is to develop the LA tool for reflective writing. To translate the theory into practice, the following steps will be undertaken: (a) the developing of a reflective writing framework (RWF) for CS; (b) the validation of this RWF by experts in the field; (c) the annotation of a dataset, using the proposed RWF; and (d) the implementation of reflective writing analysis based on supervised machine learning algorithms. This research presents the development process of an LA tool infrastructure based on the RWF designed for use in the CS arena. In particular, the research explores the framework's assessment criteria in terms of reflection indicators and levels.

2 RELATED WORK

The existing approaches for automatic reflective writing analysis are classified into keyword-based, lexical rule-based and machine learning-based categories (Alrashidi & Joy, 2020; Chng, 2018; Gibson et al., 2017; Kovanović et al., 2018; Liu et al., 2019; Shum et al., 2016; Ullmann, 2019). The keyword-based category depends on locating specific keywords, as an indication of reflection, in the input text using a keyword matching process. A list or a dictionary with various keywords refer to each text level (assumed that all automatizations are implemented with level-based models as these models are developed for assessment purposes). The presence/absence vs. frequency of the keywords can be used to analyse input text using the keyword-based approach (Ullmann, Wild, & Scott, 2012). The rule-based category depends on applying a set of rules on sentences or phrases in the text, each rule is linked to a specific reflection level (Gibson et al., 2017; Ullmann et al., 2012). An early work on machine learning on reflective writing analysis (Ullmann, 2015) using existing classification algorithms is to find patterns in each level by the pre-implemented training stage and to classify input text by the mined patterns (Kovanović et al., 2018; Liu et al., 2019; Ullmann, 2019).

2.1 Research Questions

We attempted to respond to the following research questions when developing the LA tool based on the RWF for CS education: (1) Which criteria are used for analyzing students' reflective writing in CS education? And (2) what are potential machine learning algorithms that can distinguish between reflective writing levels?

3 THE REFLECTIVE WRITING FRAMEWORK

We asked the expert to answer open-ended questions to explore the experts' perceptions and opinions on reflective writing levels and the indicators they are used. The open-ended questionnaire was developed based on standard methods (Cohen, Manion, & Morrison, 2007; Radhakrishna, 2007)

The selection of the experts who comprised a panel of experts was critical since any outcome is based on the panel members' opinions (Abou Baker El-Dib, 2007). The selection of experts was based on their breadth of academic skills in CS and their knowledge of reflection. The participant was defined as an 'expert' if they have experience of reflective writing and formative assessment, and a background in CS education. Evidence of the panel's expertise was comprised of the published books, papers and/or the teaching experiences each could exhibit. Twenty experts were invited, via email, and of these, six agreed to participate. The recommendation range from 2-10 experts, in this study for the six participants on investigating the reflective writing criteria in assessment (Gable & Wolf, 2012). The expert panels are 3 from the USA and 3 from the UK universities.

Thematic analysis is selected to be used in this study to undertake content analysis because it is one of the most straightforward ways to conduct content analysis (Braun & Clarke, 2006). The thematic analysis of the open-ended questions responses resulted in seven codes for indicators and three codes for levels of reflection; these codes we described in detail.

The expert panels only mentioned three levels of reflective, from non-reflective, reflective, and critically reflective. The analysis of the open-ended questions responses can be summarized as follows. For the indicators of the non-reflective level, two experts used "**describe**" words in their definitions of such indicators. Expert 1 stated that: "students merely **describe** what they have done or claims are made **without** any examples." Expert 3 used the word "**listing**" instead of "**describe**" when stating that "I would often see **listings** of topics to **report** covered that I would classify as non-reflective." This means that "non-reflective" texts are superficial descriptions of situations.

For the **understanding indicator**, all the experts consulted characterized the understanding level and its indicators as bordering on the reflective level. For example, Expert 5 defined this indicator as, "when students **identify** their **understanding** of competencies, we would say that reflective writing has been reached." Accordingly, the understanding indicator is considered as characterizing in both the non-reflective and the reflective levels, according to the context. For the **feeling indicator**, all the experts argued that the reflective level applies when the writer is able to identify their own thoughts and feelings. For example, Expert 3 stated that "I would look for evidence of what the students previously **thought** or **felt** on whether that had worked or not." This means that the feeling indicator in the proposed framework is related to thoughts and feelings which can be either at the reflective or at the critically reflective levels. All the experts argued that the **reasoning indicator** occurs when a writer explains a situation or issue by providing examples and causes. For example, Expert 10 stated

that “Students are able to clearly **explain** their process, what worked, what didn’t.” Expert 7 supported this point by stating that “Students **provide examples**.” Expert 3 concurred with the above, saying that, “I would look for **analysis** of problems and how they had been solved.”

For the **perspective indicator**, Expert 7 stated that this could be detected when “Students share **personal** thoughts and **connect** with **other** thoughts.” Expert 3 supported this point by saying, “Evidence of **re-evaluation** as a result of feedback from **others**.” Expert 7 and Expert 3 emphasized that perspective takes into consideration others’ perspectives. Further, the significance of the **new learning indicator** was clearly emphasized by the panel. The experts commented that they search for evidence of learning. Expert 11 said that in terms of evidence of learning, it is expected that the student shows what has been learned as “evidence of what was **learned** through reflection.” For **future action**, the panel of experts commented, that they search for the evidence of outcomes when assessing passages of reflective writing. Expert 3 expected that the student would show that they had achieved a deeper understanding of the problem that they were engaged with, as a result of producing the reflective writing “when one is able to show awareness/realization of the problems and use it **as future reference**.”

Table 1 shows all the indicators and levels of our RWF. This framework is consistent with the literature on reflective writing and on reflection theories, especially in terms of the levels defined by Wong, Kember, Chung, and Yan (1995) and the reflection indicators defined by Ullmann (2019).

Table 1 Levels and Indicators of the RWF for CS

Reflective levels	Indicators
Non- Reflective	Descriptive: the writer reports a fact from experience and/or materials
	Understanding: the writer understands and/or analyses the experience.
Reflective	Feelings: the writer identifies and/or analyses their own thoughts and feelings.
	Reasoning: the writer explains the experience by giving reasons.
Critically	Perspective: the writer shows awareness of alternatives.
Reflective	New learning: the writer integrates and/or describes new learning
	Future action: the writer intends and/or plans to do something in the future.

3.1 Proposed Work

In order to implement the automated reflective writing analysis, there is a need to develop a sophisticated mapping approach to reach the intended goal. The proposed approach for such mapping depends on intermediating this process with reflective writing indicators. There is no one-to-one mapping into the Automated Reflective Writing Analysis.

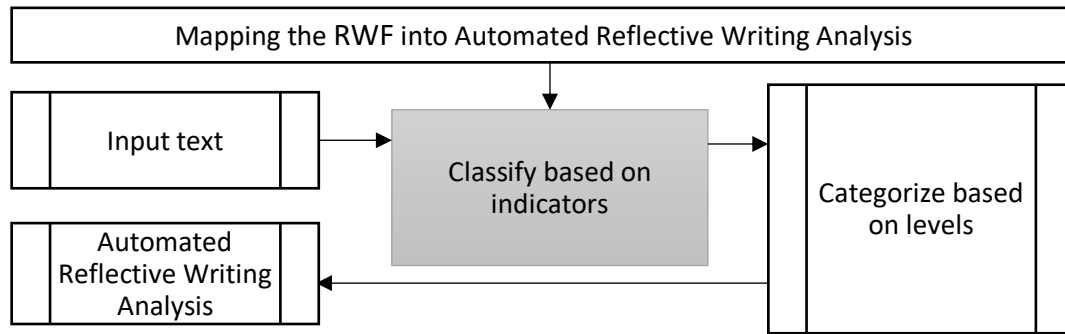


Figure 1: Mapping reflective writing indicators and levels into Automated Reflective Writing Analysis

Neither the components nor the reflection text can be mapped into the automated reflective writing analysis, which can categorize the content of the text into the automated reflective writing analysis. Nevertheless, because input text can be mapped into reflective writing indicators easily and because the reflective writing indicators have been mapped into the reflective writing levels, as similar to the Gibson, Kitto, and Bruza (2016), the linking of the analysis process has been proposed as two steps mapping as presented in Figure 1.

The one-to-one mapping cannot be achieved due to the nature of automating reflective writing detection, in which each unit of analysis (sentence, paragraph, or document) can be of a non-reflective, reflective, or critically reflective nature. the development of the automated reflective writing analysis is to analyze reflective text for the underlying problem, the 3-level framework that is combined with the multi indicators.

The proposed approach adapted from the authors to extract and use a set of features as input to a classification algorithm in order to generate a specific class or label to the input text. The extracting feature vector will use to classify the input text, using classification algorithms, into the seven indicators, the sentence can be one or more indicators. After this, the input text is classified into reflection levels (non-reflective, reflective or critically reflective) categories. The proposed implementation approach is illustrated in Fig. 2.

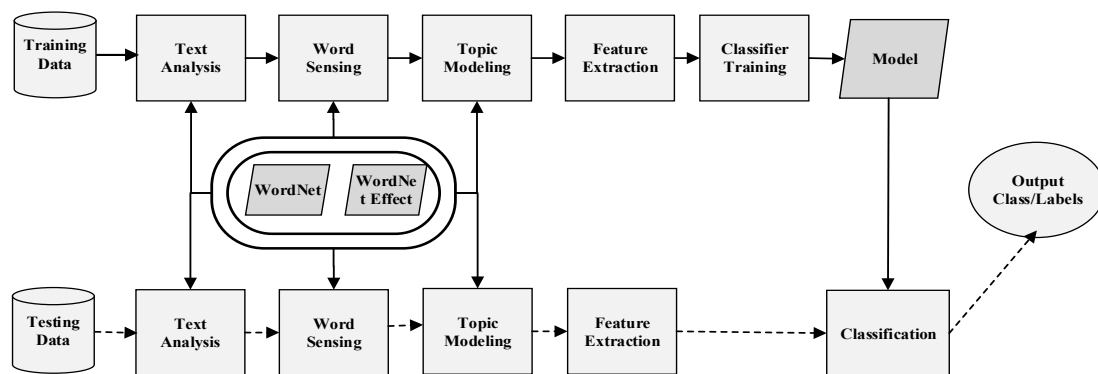


Fig. 2. The Automated Reflective Writing Analysis Approach adapted from (authors)

4 DISCUSSION

In response to the research questions, the results from the open-ended questions responses were consistent with the theoretical frameworks of reflection. The description of the non-reflective level agreed by the panel as described earlier is consistent with that in the Bain, Ballantyne, Packer, and Mills (1999) framework in which this level is described as the '**reporting**' level which is said to occur when the writer **describes, reports** or **re-tells** without added comments or insights. Hatton and Smith (1995) similarly stated that a clearly **descriptive-only** passage will include only a description of experiences that have occurred, **without** any attempt to give an **explanation** of those experiences. Ullmann (2015) highlighted the 'description of an experience' as a means of capturing the context of a piece of reflective writing - which may well be the reason that the student embarked on the reflection in the first place. This makes a **descriptive** indicator occurs when the writer reports a fact from experience and/or materials.

Birney (2012) indicated the importance of insightful understanding as evidence of reflection activities (at the understanding level): *"The student demonstrates an insightful understanding of an event or topic, e.g., a discussion of an event or understanding of that event or topic that shows a deep understanding."* That makes **understanding** indicator can be in any level of reflective writing.

the description of the **feeling** indicator in the proposed framework is consistent with the parallel descriptions in the reflection frameworks proposed in (Ullmann, 2015) – to the effect that the feelings or thoughts evinced by the experience often can be discussed for this indicator to be triggered.

As for the **reasoning** indicator, Vong (2016) noted that students display the characteristics of reasoning when they evidence thinking about the experience or when they provide in-depth interpretations of the events in question. Thus in line with our description of the **reasoning** that the writer explains the experience by giving reasons.

Our description of the **perspective** indicator is consistent with Moon (2004), who discussed perspective in terms of "evidence of external ideas or information." Ullmann (2015) described that *"the perspective of someone else, theory, the social, historical, ethical, moral, or political context."* Thus in line with our description of the **perspective** indicator that the writer shows awareness of alternatives.

The new learning indicator is described in many reflection frameworks (Moon, 2004; Prilla & Renner, 2014; Ullmann, 2015; Wong et al., 1995) using similar concepts to the one presented in the proposed framework here. Ullmann (2015) described that *"Descriptions of the lessons learned."* Thus in line with our description of the **new learning indicator** that the writer integrates and/or describes new learning.

Our description of the **future action** indicator is in the line (Birney, 2012; Ullmann, 2015) that the writer would, given the same circumstances again, intentionally do something differently or they would make a plan of action based on the new understanding that has resulted from considering and reviewing the original experience

The framework proposed here was defined by the findings of the open-ended questionnaires with expert in CS education in higher education that are consistent with the literature on reflective writing

and reflection theories especially in terms of selected frameworks and in terms of the levels defined by (Wong et al., 1995) and the reflection indicators defined by (Ullmann, 2015). In conclusion, the panels of experts clarified the levels and indicators associated with reflective writing in the computer science field. Moreover, the analysis of the qualitative criteria led to the elucidation of the relationship between the reflection indicators and the associated levels.

5 CONCLUSION

This research has answered two research questions that aimed to explore 1) the criteria used for analyzing students' reflective writing in CS education; and 2) potential machine learning algorithms that can distinguish between reflective writing levels. Based on the content analysis of the open-ended questions responses, the RWF was proposed; this has three levels and seven indicators, specifically to assess reflective writing produced in the context of CS education. Thus, we build the underpinning of the RWF to develop the LA tool of reflective writing.

We plan, in the overcoming years, to create a corpus of reflective writing in CS education in order to investigate the language and linguistic features used for reflective writing within CS. We also aim to automate the framework by designing an LA tool based on rule-based and machine learning algorithms to determine the features of reflective writing samples. This would be challenging to automate quality feedback which requires to set significant rules and annotate quality corpus. An automated assessment system would mean students could have instant feedback on areas in which they have weaknesses. Moreover, we aim to study the use of our RWF for the enhancement of the educational impact of such feedback.

ACKNOWLEDGMENT

The research was partially funded by Kuwait Foundation for the Advancement of Sciences (KFAS); Project code CB19-68SM-01; and Kind Saud bin Abdulaziz University for Health Sciences (KSAU-HS); scholarship program for Teacher Development, Department of Languages and Cultures.

REFERENCES

- Abou Baker El-Dib, M. (2007). Levels of reflection in action research. An overview and an assessment tool. *Teaching and teacher Education*, 23(1), 24-35. doi:10.1016/j.tate.2006.04.002
- Alrashidi, H., & Joy, M. (2020, 2020//). *Reflective Writing Analysis Approach Based on Semantic Concepts: An Evaluation of WordNet Affect Efficiency*. Paper presented at the Intelligent Systems and Applications, Cham.
- Bain, J. D., Ballantyne, R., Packer, J., & Mills, C. (1999). Using journal writing to enhance student teachers' reflectivity during field experience placements. *Teachers Teaching*, 5(1), 51-73.
- Birney, R. (2012). *Reflective Writing: Quantitative Assessment and Identification of Linguistic Features*. (Ph.D. Dissertation. Thesis). Waterford Institute of Technology,
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. doi:10.1191/1478088706qp063oa
- Chng, S. I. (2018). Incorporating reflection into computing classes: models and challenges. *Reflective practice*, 19(3), 358-375.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research Methods in Education*.
- Corich, S. P. (2011). *Automating the measurement of critical thinking in discussion forums: a thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy (Ph.*

- D.) in *Information Systems at Massey University, Palmerston North, New Zealand*. Massey University,
- Fekete, A., Kay, J., Kingston, J., & Wimalaratne, K. (2000). *Supporting reflection in introductory computer science*. Paper presented at the ACM SIGCSE Bulletin.
- Gable, R. K., & Wolf, M. B. (2012). *Instrument development in the affective domain: Measuring attitudes and values in corporate and school settings* (Vol. 36): Springer Science & Business Media.
- George, S. E. (2002). Learning and the reflective journal in computer science. *Australian Computer Science Communications*, 24(1), 77-86.
- Gibson, A., Aitken, A., Sándor, Á., Buckingham Shum, S., Tsingos-Lucas, C., & Knight, S. (2017). *Reflective writing analytics for actionable feedback*. Paper presented at the Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17.
- Gibson, A., Kitto, K., & Bruza, P. (2016). Towards the Discovery of Learner Metacognition from Reflective Writing. *Journal of Learning Analytics*, 3(2), 22-36. doi:10.18608/jla.2016.32.3
- Hatton, N., & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation. *Teaching and teacher Education*, 11(1), 33-49.
- Kovanović, V., Joksimović, S., Mirriahi, N., Blaine, E., Gašević, D., Siemens, G., & Dawson, S. (2018). *Understand students' self-reflections through learning analytics*. Paper presented at the Proceedings of the 8th international conference on learning analytics and knowledge.
- Liu, M., Shum, S. B., Mantzourani, E., & Lucas, C. (2019). *Evaluating Machine Learning Approaches to Classify Pharmacy Students' Reflective Statements*. Paper presented at the International Conference on Artificial Intelligence in Education.
- Moon, J. (2004). Using reflective learning to improve the impact of short courses and workshops. *Journal of Continuing Education in the Health Professions*, 24(1), 4-11.
- Moseley, D., Baumfield, V., Higgins, S., Lin, M., Miller, J., Newton, D., . . . Gregson, M. (2004). *Thinking Skill Frameworks for Post-16 Learners: An Evaluation. A Research Report for the Learning and Skills Research Centre*: ERIC.
- Poldner, E., Simons, P., Wijngaards, G., & Van der Schaaf, M. (2012). Quantitative content analysis procedures to analyse students' reflective essays: A methodological review of psychometric and edumetric aspects. *Educational Research Review*, 7(1), 19-37.
- Prilla, M., & Renner, B. (2014). *Supporting Collaborative Reflection at Work*. Paper presented at the Proceedings of the 18th International Conference on Supporting Group Work - GROUP '14.
- Radhakrishna, R. B. (2007). Tips for developing and testing questionnaires/instruments. *Journal of extension*, 45(1), 1-4.
- Ryan, M. (2011). Improving reflective writing in higher education: A social semiotic perspective. *Teaching in Higher Education*, 16(1), 99-111.
- Shum, S. B., Sándor, Á., Goldsmith, R., Wang, X., Bass, R., & McWilliams, M. (2016). *Reflecting on reflective writing analytics: Assessment challenges and iterative evaluation of a prototype tool*. Paper presented at the Proceedings of the sixth international conference on learning analytics & knowledge.
- Thorpe, K. (2004). Reflective learning journals: From concept to practice. *Reflective practice*, 5(3), 327-343.
- Ullmann, T. D. (2015). *Keywords of written reflection-a comparison between reflective and descriptive datasets*. Paper presented at the CEUR Workshop Proceedings.
- Ullmann, T. D. (2019). Automated Analysis of Reflection in Writing: Validating Machine Learning Approaches. *International Journal of Artificial Intelligence in Education*, 1-41.
- Ullmann, T. D., Wild, F., & Scott, P. (2012). *Comparing automatically detected reflective texts with human judgements*. Paper presented at the Proceedings of the 7th European Conference on Technology Enhanced Learning (EC-TEL 2012), Saarbrücken, Germany.
- Vong, S. (2016). REPORTING OR RECONSTRUCTING? The zine as a medium for reflecting on research experiences. *Communications in Information Literacy*, 10(1), 3.

- Wing, J. M. (2014). Computational thinking benefits society. *40th Anniversary Blog of Social Issues in Computing, 2014*.
- Wong, F. K., Kember, D., Chung, L. Y., & Yan, L. (1995). Assessing the level of student reflection from reflective journals. *Journal of advanced nursing, 22*(1), 48-57.

Facilitating Adaptive Assessment Delivery at Scale: Echo-Adapt Software-As-A-Service

Michelle Barrett, Bingnan Jiang

Act, Inc.

{michelle.barrett, bingnan.jiang}@act.org

ABSTRACT: Modern assessment platforms deliver demanding content to effectively estimate test takers' ability and help to guide their learning. More efficient measurement can be realized by using adaptive assessment. This type of assessment, which relies on adaptive algorithms to create near-real time tailored testing for a large group of test takers, raises new challenges for test delivery reliability, test security, and test development. First, modern adaptive assessment platforms are usually deployed to the cloud to deliver computerized adaptive testing (CAT) for good scalability. Yet, any network outage, server breakdown, or even bursts of concurrent testing requests could disrupt the ongoing adaptive assessment process. Second, since items are usually limited in the item bank, some popular items are likely to be extensively exposed to test takers and thus test security is compromised. Third, the potentially inappropriate implementation of adaptive assessment systems, e.g., configuring algorithms and models using complex scripts, could impose unnecessary obstacles to test development at scale, especially for users who are not experts on adaptive assessment or coding. In this paper, we will discuss the way that Echo-Adapt[®] overcomes these challenges as software-as-a-service. First, the shadow-test approach guarantees the reliable test delivery even when the connectivity to the cloud server is accidentally lost. Second, the ineligible constraint method dynamically controls item usages and exposure rates over a large group of test takers. Third, the development of interactive user interface allows non-CAT experts to easily configure testing and establish content specification constraints. In addition, we will discuss the approach to integrate Echo-Adapt[®] CAT APIs with IMS QTI compliant test delivery platforms and present empirical results of item exposure control and user experience.

Keywords: Adaptive assessment; computerized adaptive testing, shadow test; item exposure control; IMS QTI;

1 INTRODUCTION

As modern classrooms are increasingly driven by technologies, traditional paper-and-pencil assessments are also transforming to digital assessments that rely on advanced psychometric models, emerging technologies, and multimedia contents. Traditional fixed-form assessments are usually composed of items from a wide range of difficulty and thus inefficient to measure skills and knowledge of examinees of high or low abilities. Alternatively, computerized adaptive testing (CAT) administers tailored testing to estimate the ability of an examinee with reduced testing time, improved accuracy, increased security, and reliable delivery. To date, well-designed adaptive algorithms have been shown to produce a reasonably stable estimate of an examinee's ability within about 10 items (van der Linden & Pashley, 2010). However, CAT raises new challenges in terms of test delivery reliability, test security, and test preparation, especially for large-scale assessments. To overcome or minimize these challenges, specific design principles should be followed and combined with efficient implementations. This paper presents the design principles and solution approaches for adaptive

assessment delivery at scale along with the description of Echo-Adapt, commercial software-as-a-service (SaaS) for large-scale adaptive assessments.

2 DESIGN PRINCIPLES

A variety of design principles should be followed to build a high-performant and reliable adaptive assessment platform. This section is focused on the discussion of principles with respect to item delivery latency and reliability, test security, and CAT configurability.

2.1 Latency Optimization & Reliable Delivery

A well-designed adaptive assessment platform must guarantee consistently low latency for item delivery in all applications because examinees would be impatient and distracted if they have to wait a long time for next items to show on the screen. Use experience testing has found that a user will expect a latency of no longer than 1 second to feel like they are navigating freely (Nielsen, 2009). As latency increases beyond 1 second, users begin to suspect an issue with the system. The latency issue is usually common and critical for large-scale assessment (e.g., Hill, 2013; Tanner, 2019) and may be exacerbated by CAT deliveries. As tens of thousands of concurrent CAT sessions consume huge resources of computing, storage, and network bandwidth from the cloud, the limited resources allocated to each session can bring down the runtime performance of CAT delivery to an unacceptable level. A major part of item delivery latency is from the CAT algorithm runtime, which should be optimized in the design and implementation of CAT platforms.

A worse scenario is the interruption of CAT delivery caused by server outages or network congestions. As CAT delivers items sequentially, this could cause the termination of a test administration. Thus, the design of a CAT platform should support the continuity of item delivery even without connection to the cloud environment.

2.2 Test Security

An adaptive assessment system intends to deliver the optimal item (e.g., with maximum fisher information at the current ability estimate) to an examinee at each stage. It is very likely that some popular items are excessively delivered and exposed to examinees so that their security is compromised. It is necessary to balance items/passages usage for test security and prevent potential cheating. A workaround is to select a sub-optimal item if the optimal one has already been over-exposed. Empirical results show that the accuracy of the ability estimates in adaptive assessments is not sacrificed too much by adopting the exposure control strategy (van der Linden & Veldkamp, 2007).

2.3 Configurability

The test configuration for a large-scale adaptive assessment delivery is usually complex, with content specification constraints at multiple levels and various algorithm parameters to tune. This requires multiple rounds of simulations and modifications on the configuration before the live administration. Usually, multiple users who operate with different subject matter, e.g., psychometricians and content developers, work together to edit the same test configuration for the assessment delivery. These users are of different background. For example, some lack coding skills while others don't know much about the psychometric models. Therefore, it is necessary to separate the configuration of an algorithm from

the algorithmic coding itself. This allows users who are not software engineers to design, configure, and deploy an adaptive assessment delivery. It also makes the configuration process more robust.

3 SOLUTION APPROACHES

In this section, we describe the solution approaches to meet the design principles discussed in Section 2. As a well-recognized approach to CAT, the shadow-test approach (van der Linden, 2005) solves the dilemma by assembling full-length shadow tests as a part of selecting items throughout the testing process. Shadow-test CAT has many advantages, including the full coverage of the test blueprint, separation of test specifications from CAT algorithms for easily modifiable configurations, and support for flexible and reliable delivery options. Therefore, the shadow-test approach is selected as the framework to incorporate other psychometric and statistical models.

3.1 Shadow-Test Approach to Optimal and Reliable CAT Delivery

The shadow-test approach sequentially assembles full-length test forms (shadow tests) based on the real-time update of the examinee's ability estimate. Shadow-test assembly is modeled as a mixed integer programming (MIP) problem. A MIP optimizes (either in the minimization or maximization sense) a function of variables (the objective) by selecting the best possible set of decisions (Smith & Taşkın, 2007). A typical shadow-test assembly MIP selects a subset of items from an item pool to maximize the test information

$$\arg \max f(x_{ij}) := \sum_{i_j \in S} I_{i_j}(\hat{\theta}) x_{i_j} \quad (1)$$

Subject to content specification constraints

where S is the set of items in the item pool, $I_{i_j}(\hat{\theta})$ is the Fisher information of item i associated with passage j at the examinee's ability estimate $\hat{\theta}$, and x_{i_j} is the binary decision variable for the selection of item i_j in the shadow test. $x_{i_j} = 1$ if item i_j is selected in the shadow test, otherwise $x_{i_j} = 0$. The shadow test MIP model defines content specification constraints at different levels for items, passages/stimuli, and the entire test. A non-exhaustive set of examples of content specification constraints include: 1) test length, 2) enemy items, 3) average attributes of selected item/passages, and 4) passage positions in the test.

At the beginning of an adaptive stage, the shadow-test approach administers optimal items in two steps as shown in Figure 1. The first step is to build the shadow test by solving the shadow-test assembly MIP based on the updated examinee's ability. It selects a set of optimal items from the item pool based on the examinee's ability estimate while conforming to all constraints. A shadow test consists of two parts, a set of items that have already been administered and a set of items that are unseen to the examinee. The second step is to administer the optimal item with maximum information from the set of unseen items while conforming to rules like correct passage order in test and correct item order in a passage. When a shadow test is assembled for the next adaptive stage, all previously administered items are constrained to be selected in the MIP model.

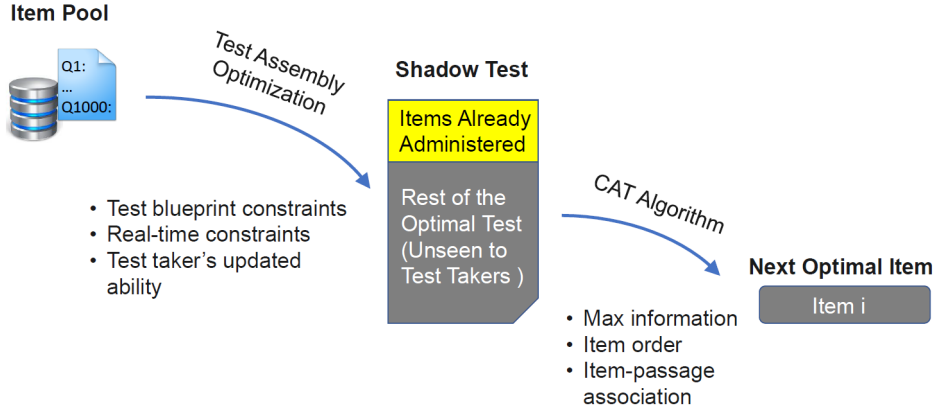


Figure 1: Item administration through the shadow-test approach

The shadow-test approach also prevents interruption of CAT delivery when server outage or network congestion happens. This is because a shadow test consists of all items to complete the entire CAT delivery. Suppose there is a sever or network failure at stage k and the CAT delivery platform cannot receive the updated shadow test from the cloud service before timeout. Since the CAT delivery platform has already received the full-length shadow test at the previous stage $k - 1$, it continues to deliver the next item from the unseen part of shadow test. Although the item delivery may not be optimal in this case, the shadow-test approach guarantees the continuity of CAT delivery and minimizes the impact from unexpected outage. Once the server is back online, the process of sequentially assembling shadow tests will be resumed to deliver optimal items for the remaining stages.

3.2 Item Exposure Rate Control

An effective method to control item exposure rate in the shadow-test approach is the ineligible constraint method (van der Linden & Veldkamp, 2007). This method describes item/passage administration eligibilities in an $I \times K$ probability matrix, where I is the number of items/passages in the pool and K is the number of contiguous intervals across the theta continuum (from $-\infty$ to $+\infty$). The probability $\hat{P}^{(j+1)}(E_i|\theta_k)$ is calculated to determine if an item/passage i is eligible for administration to an examinee with ability in the theta range k

$$\hat{P}^{(j+1)}(E_i|\theta_k) = \min \left\{ \frac{r_{\max} \varepsilon_{ijk}}{\alpha_{ijk}}, 1 \right\}, \text{ for } \alpha_{ijk} > 0 \quad (2)$$

where r_{\max} is the exposure goal rate, α_{ijk} is the number of examinees through examinee j who visited theta range k and took item/passage i , and ε_{ijk} is the number of examinees through examinee j who visited theta range k when item/passage i was eligible. For item i that has not been administered, i.e., $\alpha_{ijk} = 0$, the related $\hat{P}^{(j+1)}(E_i|\theta_k) = 1$. The eligibility probabilities are then used to conduct $I \times K$ binomial experiments

$$X_{ik} \sim B(1, p) \quad (3)$$

where $p = \hat{P}(E_i|\theta_k)$. If $X_{ik} = 0$ then item/passage i is ineligible at theta interval k ; otherwise the item/passage is eligible. To avoid the infeasibility (no solutions) of shadow-test assembly, the item/stimulus ineligibility constraints are added to the MIP model as soft constraints. Specifically, a penalty term is subtracted from the objective function when ineligible items/stimuli are selected in the shadow-test

$$\text{Maximize} \quad \sum_{i_j \in S} I_{i_j}(\hat{\theta})x_{i_j} - M \sum_{i_j \in V} x_{i_j} \quad (4)$$

where V is the set of ineligible items based on the results of exposure control experiment. The penalty M is selected as a value greater than the maximum item information value of the items in the pool at the current ability estimate. The penalty term avoids selecting ineligible items if feasible shadow tests can be assembled after excluding them, because the selection of infeasible items will decrease the MIP objective value that is to be maximized. Otherwise, ineligible items are still allowed for selection to prevent infeasibility and test interruption.

3.3 Intuitive User Interface with Smart Feedback

Instead of writing complex and error-prone code to configure CAT and specify content constraints, users should be provided an intuitive user interface (UI) to complete all the jobs through clicking and dragging. A UI informs the users with the configurable parameters of an algorithm and their valid numeric ranges. To enhance system stability, A UI should hide specific parameters since modifying them could cause unexpected performance issues or even system crashes. To further assist users with configurations, a well-designed UI should also display smart feedback information as the response to users' actions and inputs. For example, when an invalid parameter is specified by a user, the UI should show an error message that also includes the valid numeric range to fix the issue.

4 ECHO-ADAPT SOFTWARE-AS-A-SERVICE

In this section, we describe Echo-Adapt (version 1.58), the commercial SaaS to deliver adaptive assessments at scale. It follows the design principles in Section 2 and implements the solution approaches in Section 3 to deliver personalized testing with maximum efficiency, reliability, and satisfying user experience. We focus on the discussion of its software architecture, high-level design, API integration, and user interface.

4.1 Overview of Architecture & Implementation

The Echo-Adapt software architecture and its interaction with users and other platforms are shown in Figure 2. Echo-Adapt consists of three major components, including the intuitive UI, RDS database, and CAT engine. All components are deployed on Amazon Web Services (AWS) and loosely coupled by APIs to conduct CAT related tasks.

Users of Echo-Adapt are content developers and psychometricians who can create and tune CAT configurations on the same UI for live CAT administrations. Different users can access the same CAT configurations and conduct tasks including uploading item pools, adding/modifying content specification constraints, setting algorithm parameters, and running simulations. Please note that

Echo-Adapt is not seen by examinees. Although examinees are presented with Echo-Adapt's item choices, the item contents are presented by the test delivery platform.

The CAT configuration, as well as uploaded item pool data and interim CAT data, are persisted in the AWS RDS database. The Echo-Adapt CAT engine retrieves configuration data from RDS via API calls and conducts CAT tasks in parallel, i.e., live item administrations for CAT, CAT simulations, and configuration feasibility check. The computing capacity is pre-scaled based on the potential peak demand of CAT tasks to meet required performance, i.e., less than 500ms latency for an item administration given 40,000 concurrent examinees. Thus, the CAT engine is deployed on multiple Amazon Elastic Compute Cloud (Amazon EC2) instances to enhance system reliability and balance runtime performance and operational cost for large-scale assessments. In live CAT administrations, Echo-Adapt communicates with the test delivery platform via APIs that comply with the IMS Global Question & Test Interoperability (QTI) specification (IMS Global, 2015). The information exchange includes selected item identifiers and scores. Echo-Adapt does not share item content with the test delivery platform because it is directly handled by the latter.

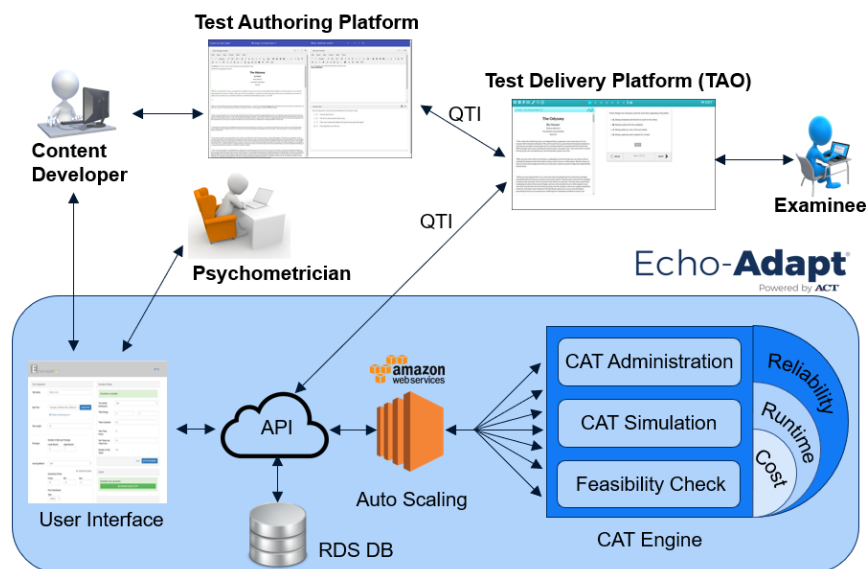


Figure 2: Echo-Adapt SaaS architecture

The Echo-Adapt high-level design diagram is shown in Figure 3, which describes the system goals, enabling technologies, and psychometric/statistical models. Echo-Adapt is designed for large-scale adaptive assessments and can be easily integrated with any IMS QTI compliant test delivery platforms. The design also follows the principles to enable the intuitive configuration process. To guarantee the system performance and reliability, Echo-Adapt is implemented using many advanced technologies: 1) The CAT engine and UI are implemented in Java and JavaScript, respectively. 2) The shadow-test MIP is implemented in Mosel scripting language and solved by a high-performant commercial MIP solver. 3) All models and components are deployed and run on the AWS cloud environment that can be easily scaled out for large-scale assessments. With respect to psychometric models, Echo-Adapt chooses the 3-parameter logistic item response theory (3PL IRT) model as the response model that is embedded in the shadow-test model. Additional statistics models are implemented to complete the item delivery and ability estimate for CAT, e.g., post-MIP processing and scoring methods.

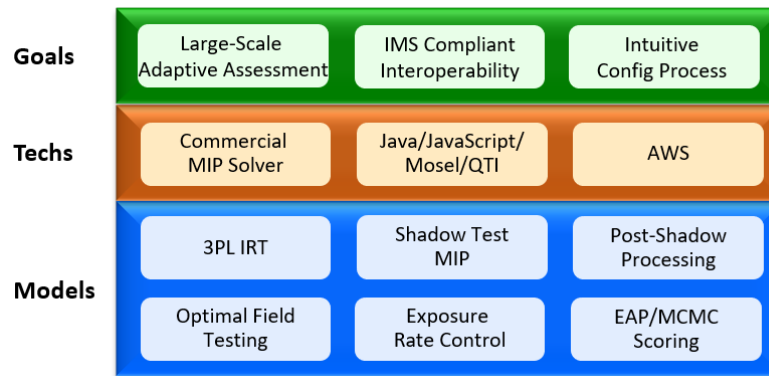


Figure 3: Echo-Adapt high-level design diagram

4.2 API Integration

Interoperability is a critical implementation issue that needs to be addressed in any assessment platforms. Establishing good interoperability between adaptive testing engines and test delivery platforms eliminates the requirement for costly proprietary integrations. Echo-Adapt is designed and built to conform to the IMS Global QTI standards to ensure a high degree of interoperability with test delivery platforms.

QTI standards for CAT define a format for the exchange of test questions to deliver, scoring information for individual questions, the examinee's interim and final ability estimates, precision of the ability estimates, etc. Multiple options exist for architecture of the test delivery platform and the adaptive testing engine; an adaptive engine delivered as SaaS is typically implemented as a service and usually accessed by the test delivery platform using the HTTPS protocol. QTI compliant CAT engine APIs define actions including creating test session, verifying items, submitting results, ending a test for an individual examinee, and ending test session. Their implementations follow the OpenAPI specification (formally known as Swagger Specification). An example of HTTP request from a test delivery platform to submit results for the first item is shown in Figure 4. The response from Echo-Adapt to the test delivery platform includes not only the item identifier for the next stage but also identifiers for the other remaining stages to guarantee the continuity of delivery with server outage.

```
POST BASE_URL/sections/632/sessions/04a87995-1090-433d-9bb6-a5f9378b3dee/results
```

```
1. {  
2.   "assessmentResult":{  
3.     "itemResult":[  
4.       {  
5.         "identifier":"T105_32418", ← Item Administered  
6.         "outcomeVariables":[  
7.           {  
8.             "identifier":"SCORE",  
9.             "value":[  
10.              {  
11.                "baseType":"integer",  
12.                "value":1 ← Item Score  
13.              }  
14.            ]  
15.          }  
16.        ]  
17.      }  
18.    ],  
19.  },  
20.  "sessionState":{"eyJhZGFwZGdlZXN0YXdIdjoxLDCJWcmV2aw9lc2x5UzV2Ibk1OZWl1IjpbXSwibmV4dE10ZWl1IjpbIkxxMDVfMzI0MTgicCJJMTA1XzMyNDE0IiwISTEWNvNyNjU1MiJDLCAzaGk6b3duZXN0IjpbIkxxMDVfMzI0MTgicCJJMTA1XzMyNDE0IiwISTEWNvNyNjU1MiJDLCApdGVtU2NmcmVzIjpbXSwiZWNUaGV0YVBvaW50cyIE6bnVsbcWiZWNFbGlnaWJpbG10eUluZGl1YyJyXvcnMiOm51bGwsImVjVHlwZSI6MSwiOmVmXndWvdE1bWJ1ciE6MiwidGZvdFRhaZVyuZVzc2I2Vkbk1IjoimDRhOdC50UTuMTA5MC0zMnk1LTliYmItYTVMOTM3OGIzIGV1In0="}
```

Figure 4: Request from a test delivery platform to submit results for the first item

4.3 User Interface & Process of CAT Configuration

Configuring a CAT in Echo-Adapt is simple and intuitive. All steps can be done on the UI without coding requirement and software redeployment. The screenshot of the CAT configuration UI is shown in Figure 5. The typical process of configuring a CAT and running simulation is shown in Figure 6. The first step is to upload load item and passage data. Echo-Adapt accepts item and passage pools in zipped csv files. The csv files include item/passage statistics and categorical attributes. As a type of smart feedback, Echo-Adapt validates the item/passage data format when they are uploaded. If the format validation fails to pass, then Echo-Adapt displays not only the error message but also the suggestions to fix the issues. After the item/passage data are uploaded successfully, the next step is to specify content specifications. Users need to specify the test length, number of passages, and add content constraints using the constraint editor. Whenever users save a modified configuration, they can choose to validate the content specifications, e.g., to check if there are any conflicting constraints, and evaluate the strictness of these constraints based on the returned number of feasible forms. If the content specifications pass the feasibility validation, users can continue to configure parameters of CAT algorithms and run simulations. Simulation results can be downloaded in a csv file for analysis and visualization. If simulation results meet the CAT design requirement, the CAT configuration can be finalized (locked for editing) for the live administration.

The screenshot displays the Echo-Adapt user interface for CAT configuration. The header includes the Echo-Adapt logo (Powered by ACT) and links for Help and Log Out. The main content area is divided into sections for configuration:

- Test Configuration** (selected tab) and **CAT** (inactive tab).
- Test Name**: A text input field.
- Item Pool**: A dropdown menu with an "Upload New" button and a "Display item pool" link.
- Test Length**: A text input field.
- Enable Enemy Items**: Radio buttons for "Yes" and "No".
- Test Configuration Type**: Radio buttons for "ATA" and "CAT".
- Exposure Control**:
 - Inherit Exposure Control Data**: A dropdown menu with "From" and "None" options.
 - Exposure Control Type**: A dropdown menu with "None" as the selected option.
 - Randomize the First L Items**: A text input field with "L=" and "0" as the value.
- Scoring Method**: A dropdown menu with "EAP" as the selected option. An "Advanced Options" link is visible.
- Constraints (0)**: A section with a "New Constraint" link.
- Feasibility Analysis Cutoff**: A text input field with "1000" as the value.
- Total Feasible Tests**: A section with the text "Results calculated upon saving configuration."

 At the bottom right, there are "Reset" and "Save Configuration" buttons.

Figure 5: Echo-Adapt user interface for CAT configuration

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

In a short survey of current users of Echo-Adapt, the majority of users are satisfied with the design of UI and the process of CAT configuration. Users especially like the design of constraint editor and how easily it can be used to add various content specification constraints. Some users also vote positive for the ability to add constraints on the fly and review the strictness of test blueprint through iterations.

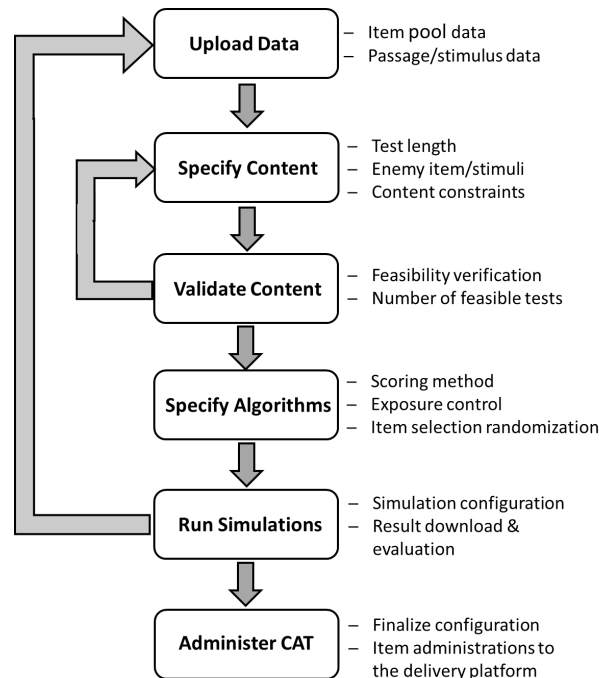


Figure 6: Process of configuring and administering CAT in Echo-Adapt

5 DISCUSSION

This paper presents the design principals and solution approaches that are essential to overcome the new challenges for the adaptive assessment delivery at scale. It has been focused on the critical issues in terms of latency, reliability, security, and configurability. In addition, the design and implementation of Echo-Adapt, commercial SaaS for adaptive assessment delivery at scale, are presented as an example of applying the principals and approaches to the real-world application for the operational CAT delivery. The intuitive user interface and easy process of configuration in Echo-Adapt demonstrate its simplified while robust configurability for adaptive assessment delivery at scale.

REFERENCES

- Echo-Adapt (Version 1.58). [Computer Software]. Iowa City, IA: ACT.
- IMS Global (2015). *IMS question & test interoperability (QTI) specification*. Retrieved from <http://www.imsglobal.org/question/index.html>
- Hill, R. (2013). An analysis of the impact of interruptions on the 2013 administration of the Indiana Statewide Testing for Educational Progress—Plus (ISTEP+). Retrieved February 2020 from <https://www.nciea.org/sites/default/files/publications/ISTEP%20RH072713.pdf>.
- Nielsen, J. (2009). Powers of 10: Time Scales in User Experience. Retrieved April 2017 from <https://www.ngroup.com/articles/powers-of-10-time-scales-in-ux/>.

- Smith, J. C., & Taşkın, Z. C. (2007). A Tutorial Guide to Mixed-Integer Programming Models and Solution Techniques. Retrieved December 6, 2018 from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.464.6182&rep=rep1&type=pdf>
- Tanner, C. (2019). Computer glitch stops 18,000 Utah students from finishing end-of-year tests. Retrieved February 5, 2020 from <https://www.sltrib.com/news/education/2019/04/26/computer-glitch-stops/>.
- van der Linden, W. J. (2005). *Linear models for optimal test design*. New York, NY: Springer.
- van der Linden, W. J., & Pashley, P. J. (2010). Item Selection and Ability Estimation in Adaptive Testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Elements of Adaptive Testing* (pp. 3–30). New York, NY: Springer New York. https://doi.org/10.1007/978-0-387-85461-8_1
- van der Linden, W. J., & Veldkamp, B. P. (2007). Conditional Item-Exposure Control in Adaptive Testing Using Item-Ineligibility Probabilities. *Journal of Educational and Behavioral Statistics*, 32(4), 398–418. <https://doi.org/10.3102/1076998606298044>

Providing Directed Feedback Through QUICK-Comments

Anthony F. Botelho, John A. Erickson, Aaron G. Alphonsus, Neil T. Heffernan

Worcester Polytechnic Institute, Worcester MA

abotelho@wpi.edu; jaerickson@wpi.edu; agalphonsus@wpi.edu; nth@wpi.edu

ABSTRACT: Online learning platforms provide opportunities to deliver aid and feedback to students, but also provide a means to augment a teacher’s ability to attend to student needs. Despite a growing focus on incorporating personalization into the functions of a learning system, many such efforts exclude or ignore the role of teachers. Particularly in the context of assigning open-ended questions, such as those that are becoming more common in mathematics open educational resources, teachers spend hours attempting to provide direct and meaningful feedback to their students. In this paper, we describe the development of a tool called QUICK-Comments that, deployed through an online learning platform, attempts to augment teachers’ ability to provide personalized, meaningful feedback to students answering open-ended questions on middle school mathematics topics. By leveraging machine learning, natural language processing, and teacher-in-the-loop development, we describe an overview of the motivations behind the tool and address the open challenges that are being faced as we begin pilot testing QUICK-Comments.

Keywords: Feedback; Natural Language Processing; Learning Systems; Machine Learning

1 INTRODUCTION

The adoption of learning platforms in classrooms across all grade levels provides teachers with better tools to assess student knowledge and augment their ability to provide feedback and instruction to students. In many modern systems, it is not uncommon for such learning platforms to strive for a level of personalization in the types of aid or delivery of content to students. However, it is important to truly understand what it means to “personalize” support for students. On the surface, the goal of personalization may appear rather intuitive; a “personalized” learning system aims to provide each student with content, feedback, and instruction that maximizes learning for the individual. This differs from a traditional, technology-free classroom where the delivery of content, feedback, and instruction is typically selected by the teacher to benefit to the class as a whole.

Many visions of providing personalized instruction to students, however, exclude or ignore the role of the teacher. While often difficult, teachers are often very good at providing directed feedback and instruction to students in need of help. Considering the numerous responsibilities that a teacher must fulfill (e.g., constructing lesson plans and designing instruction), as well as the number of students in an average classroom, it is simply infeasible for a teacher to sufficiently attend to every student with directed feedback and instruction; as will be addressed in Section 1.2, this is particularly true in the case of responding to student answers to open-ended questions.

How can we, as researchers and developers of educational technologies, build the right set of tools to support teachers in providing meaningful, personalized feedback to benefit their students? The goal of this paper is to address this question through the development of a tool called QUICK-

Comments, aimed at leveraging the strengths of online systems, machine learning, and natural language processing to help provide feedback for student answers to open-ended problems in a teacher-in-the-loop design. This paper focuses on the motivation behind the tool, findings and accomplishments to date, as well as some challenges that remain as the QUICK-Comments tool begins its initial stage of pilot testing.

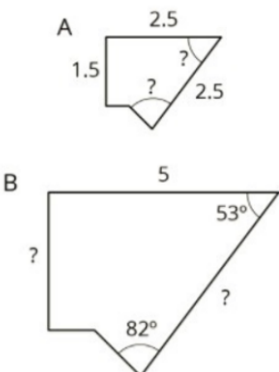
1.1 Personalization through Heterogeneous Treatment Effects

Although the ideal case of personalization is easy to define, implementation proves more difficult. In terms of causal inference, traditional methods follow a paradigm that aligns to selecting instructional policies that lead to positive average treatment effects; in other words, the delivery of content, feedback, or instruction (i.e., the treatment) is selected to help the class of students as a whole, and ignores the case that not all students may benefit in terms of an observed learning outcome. Alternatively, a paradigm that measures heterogeneous treatment effects, understanding that certain types of content, feedback, or instruction may lead to better learning outcomes for different sub-groups of students, is a way of moving toward utilizing methods that are more personalized.

This distinction is highlighted by the work of Pashler et al. (2008), who posit that in order to personalize, there must exist a specific type of subject-treatment interaction. Essentially, if

comparing one instructional policy, A, to another, B, personalization can only be achieved if it is found that A works reliably better for one sub-group of students and B works reliably better for another sub-group of students. The reason for this, of course, is that if A is found to lead to equal or greater learning gains when compared to B for all identified sub-groups of students, then there is no reason to consider instructional policy B.

A) Polygon B is a scaled copy of Polygon A.
What is the scale factor from Polygon A to Polygon B?



copied for free from openupresources.org

B) Explain your reasoning.

Figure 1: An example of the type of open response problem common to many open educational resources.

1.2 Providing Directed Feedback for Student Work

While such heterogeneous effects have been explored in a small number of previous works (c.f. Razzaq & Heffernan, 2009; Yin et al., 2017), this project considers how online systems may best aid teachers in a task where they are already providing personalized feedback to students. The use of online and computer-based platforms, particularly in the domain of mathematics, as is the focus of this paper, provide many advantages as such

content is often closed-form. In other words, the types of problems commonly observed in mathematics exhibit a single or small number of well-defined accepted answers. Computers are proficient in automating parts of student assessment for such closed-form answers. In some cases, when combined with teacher-provided content, learning systems may even be able to automatically supply students with feedback for recognized common wrong answers (Selent & Heffernan, 2014). For example, if a student is presented with a simple closed-form problem such as “What is 2×7 ?,” a system could easily provide a student who responds with the answer of “9” with a feedback message such as “Remember that ‘x’ means multiplication, not addition” Although this response is not personalized to the student, students are able to receive feedback specific to the response given.

While computers are well-suited to augment the delivery of student aid in closed-form problems, as in the previous example, the task of assessing and providing feedback to open-ended student answers is still primarily done by teachers manually. Many curricula, including widely-adopted open educational resources (OERs) such as EngageNY and Illustrative Mathematics, incorporate open-ended questions as a means of assessing each student’s understanding of the given content and ability to articulate how a solution to a given problem is reached; these problems often take the form of asking a closed-form problem, followed by an open-ended question prompting the student to explain the reasoning for, or strategy to reach, the given answer; this type of problem is exemplified in Figure 1. Through such open-ended problems, teachers have the opportunity to both assess and provide meaningful, directed feedback to identify and address gaps in student understanding or knowledge. However, as is explored in Erickson et al. (2020), teachers only score approximately 10% of student answers to assigned open response questions and provide directed feedback to approximately 2% of such answers (as illustrated in Figure 2).

The task of manually assessing and writing feedback for student answers to open-ended questions is evidently difficult for teachers to perform in tandem with other teaching responsibilities. In recognition of this problem, better tools are needed to support teachers in attending to and responding to student work. By leveraging the capabilities of online systems to augment teachers’ ability to provide directed feedback, students will be able to benefit from personalized feedback made possible through a teacher-in-the-loop set of tools.

2 QUICK-COMMENTS

In the development of tools to support teachers in assessing and responding to student answers to open-ended problems, inspiration can be drawn from many areas in- and outside of the field of education. Inspired by the manner in which Google’s SmartReply (Kannan et al., 2016) technology

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

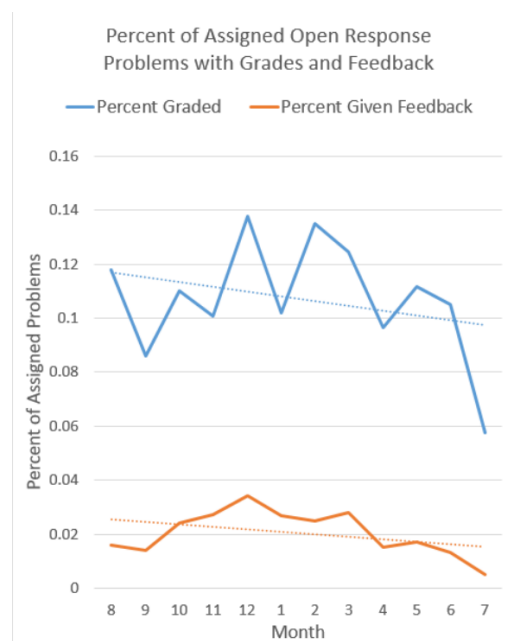


Figure 2: The percent of answers to assigned open response questions that are scored and given feedback by teachers as explored in Erickson et al. (2020)

Student ↑↓	Response ↑↓	Score ● ↑↓	Suggested Messages ●			Teacher Feedback ↑↓
1 Student	you find the slope of dividing the vertical side by the horizontal side and we already did that so the slope is 0.5	4 / 4	:)	Great job!	Perfect!	Select a suggested message or write your own comment here.
2 Student	i counted the length & width of the smaller triangle	2 / 4	Incorrect math.	Along the right lines, missing some steps	Be specific!	Select a suggested message or write your own comment here.
3 Student	3 divided by 6 is 0.5.	2 / 4	Show or explain your work.	Be specific!	This response lacked any clarity or support.	Select a suggested message or write your own comment here.
4 Student	it goes up o over 5	2 / 4	Be specific!	Along the right lines, missing some steps	Show or explain your work.	Select a suggested message or write your own comment here.

Figure 3: The user interface of the pilot version of QUICK-Comments

aids users in reading and responding to email, we are in the process of developing a tool called QUICK-Comments. Now widely adopted in many commercial settings and applications, the ability to suggest how to respond to email, notifications, or other types of messages can reduce the time required for a user to respond. The idea behind this commercial technology is being incorporated into the QUICK-Comments tool to benefit teachers and students.

While still undergoing iterative development through pilot testing, the current version of the QUICK-Comments interface is illustrated in Figure 3. Similar to the design of Google’s SmartReply, teachers are presented with three suggested “starter” feedback messages; at this stage, the suggestions are designed to help teachers start a message (that populates in the column denoted as “Teacher Feedback”) rather than serve as the entire feedback given to the student. This decision will be discussed further in Section 4. In addition, a suggested score is presented (pre-populated as a placeholder value in the Score column) in an effort to help teachers interpret their students’ performances.

Over the next few sections, we will describe the development of QUICK-Comments as well as the steps we are taking to support more directed feedback in alignment with the ultimate purpose of the tool.

2.1 ASSISTments

The development of the QUICK-Comments tool has been made possible for integration with the ASSISTments online learning platform through funding by NSF, Schmidt Futures, and other philanthropy. With users comprised primarily of middle school teachers and students, ASSISTments has become one of the few scientifically tested and proven systems to benefit student learning based on an efficacy trial run in the state of Maine (Roschelle et al., 2016). Based around the idea of providing immediate feedback to students, as well as supporting data-driven instruction for teachers, the system is used by thousands of real students each day for primarily mathematics homework and classwork. While the system does support other age groups and content domains,

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

ASSISTments is considered an ideal system for the development of the QUICK-Comments tool for its support of open-ended problems (many systems do not support such problem types in any capacity) as well as for its incorporation of OERs into its content base; by focusing OERs specifically, the potential impact of a tool such as QUICK-Comments could be far-reaching thanks to the wide usage of such resources and curricula.

Prior to the development of the QUICK-Comments tool, like many platforms that do support open-ended problems, teachers would be required to manually assess students and leave feedback in ASSISTments as previously described; the data in Figure 2 was collected from ASSISTments and includes teacher usage of such OER content.

2.2 QUICK-Comments Development

The development of a tool such as QUICK-Comments, beyond the obvious software development needs, requires a combination of machine learning, natural language processing, and, most importantly, collaboration with teachers. In order for a tool such as this to work well, it is important to truly understand the needs of teachers as well as what would be most beneficial to students. In order to develop meaningful, directed feedback, we needed to know how teachers currently/actually assess and respond to student work in practice. To aid in development, we recruited 14 middle school mathematics teachers using OERs in their classrooms to assign, score, and provide feedback for open-response problems from their respective curricula. From this data, we have not only started to develop and implement machine learning models for use in the tool, but we have also gained a better understanding of what teachers believe constitutes meaningful feedback for their students.

For example, in the data collected from our teachers, we were surprised to find very few repeated instances of feedback provided to different students. Even when students seemingly provided similar answers, teachers often applied effort to personalize and direct feedback to each individual student. While partially anecdotal at this stage, this observation, combined with teacher surveys, reveal that there is a desire to provide personalized feedback to students and there is value in doing

Grade	Example Responses
5	Because B is 2x biggest than A
4	I didn't understand ?
5	Because 2/12 time 2 equals 5
5	2.5 x 2=5
5	2.5 times 2 is 5 so the scale factor is 2 oops that is what I meant
5	Cause 2.5 divided by 5 is 0.5
3	Because the top one is 2.5 and 1.5 goes to 2.5
1	I guessed
3	Because the part on a is half the size of the one on part b.
5	2.5 times 2 is 5.
5	A has 2.5 on top and B has 5_2.5 x2is withc means that it was 2
2	I said that because two of them are equal

Figure 4: Example student answers with the corresponding teacher-provided score on a scale of 1 to 5. This example is appropriated from Erickson et al. (2020).

so; however, there is also a need for better support in doing so as well as a desire among teachers to remain actively involved in their students' learning.

With a source of data in the form of student answers paired with teacher scores and feedback messages, the underlying engine that powers the QUICK-Comments tool can be partitioned into two primary components. The first is a set of automated scoring models designed to interpret student answers and predict how a teacher would assess such student answers. The second component is a set of models designed to consider the assessed student answers and suggest feedback messages that a teacher would want to give in response to such answers. These two components are described over the next two sections.

3 DEVELOPING AUTOMATED SCORING MODELS

While a detailed description of the models designed to interpret student answers and predict an assessment score is provided in Erickson et al. (2020), this component of the tool utilizes a combination of natural language processing techniques and machine learning.

While natural language processing (i.e. the set of techniques and methods developed to help computers interpret and understand the semantics of written text) has been explored in many domains, applying such methods with mathematics open responses posed an early challenge for the development of the tool. This is particularly the case due to the incorporation of numbers, operators, and equations that are often integrated alongside traditional language within student responses; this type of data differs from more traditional natural language processing tasks in contexts such as English essay scoring. Examples of student responses along with the teacher-provided score (on a scale of 1 to 5) is illustrated in Figure 4.

With this data, the primary task of the natural language processing methods is to divide each answer into individual words (a process known as tokenization), with our specific implementation (Manning et al., 2014) considering words such as "didn't" as two separate words of "did" and "n't" to capture the difference in meaning of each part. From this point, a technique known as TF-IDF (Term Frequency-Inverse Document Frequency) (Ramos, 2003) is used to weigh each word based on an estimated value of importance (discounting common words and emphasizing words that are more representative of each answer). Words and their importance values are then used to construct machine learning models of varying complexities to predict the score provided by the teacher. Following the analyses conducted in Erickson et al. (2020), we have since started ensembling sets of simple models with more complex deep learning models that utilize more advanced natural language processing techniques such as word embeddings; the use of word embeddings, such as those supplied by a technique known as GloVe (Pennington et al., 2014), help capture the semantic meaning of words in each student's answer to make a more informed assessment.

In regard to developing the automated scoring models, the process is a well-defined supervised learning task, not unlike many other machine learning tasks seen in education, such as predicting next problem correctness (Corbett & Anderson, 1994; Piech et al., 2015) or student behavior (Botelho et al., 2017; Paquette et al., 2015); in such cases there is a clear label (the score) that is trying to be predicted from a set of inputs (student answers). However, while a score is arguably a type of directed feedback, it certainly is not the intended personalization that the tool is developed

for. The ultimate goal is to be able to suggest meaningful feedback messages that a teacher would want to send to a student. In that regard, the generation of feedback messages is a more difficult task.

4 SUGGESTING DIRECTED FEEDBACK FOR TEACHERS

In order to understand the development process of the second component of QUICK-Comments' engine, it is helpful to reiterate our intended goal of the tool. While we have already described several times that the goal is to help teachers provide directed feedback, it is important to stress that the goal of the tool is to save teachers time as they provide meaningful feedback that is personalized to each student answer. That said, the ongoing development process of this second component has initially focused on the time-saving aspect of that goal. As shown in Figure 3, the messages suggested by the tool in its pilot version are far from meaningful or direct. As a baseline, the initial set of generation models are designed to simply save teachers time by suggesting the most common messages that teachers have previously given to students exhibiting the corresponding grade. If the scoring model believes the student would be given a score of 100%, for example, the comment model may recommend starting a feedback message with "good job!" or "perfect!" as these were the most common messages given to such answers from teachers over all previous examples.

In this way, the model is more likely to help teachers initiate a more directed feedback message. A teacher who does select a suggested message to then edit or send directly helps inform the underlying machine learning model that its predictions were good; conversely a teacher who writes their own feedback message without selecting any of the suggestions helps the model learn what not to suggest. This feedback loop is important as we iteratively improve the underlying models.

But how do we take the next step to generate more meaningful, directed feedback messages? For this ongoing challenge, we need to understand how teachers group student responses as well as

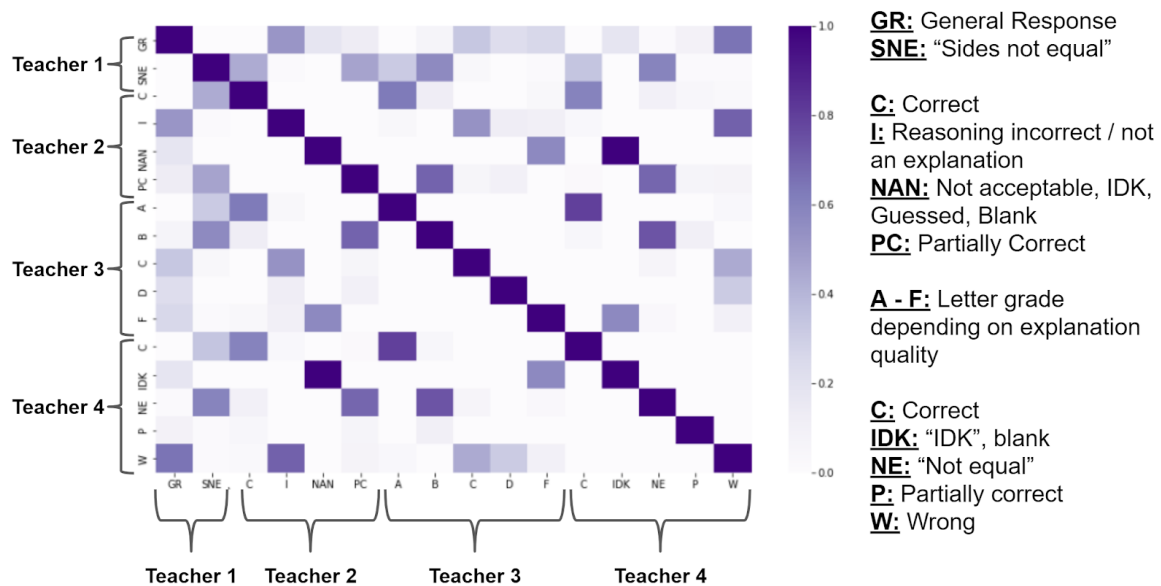


Figure 5: Example heatmap generated from 4 teachers grouping the same set of student responses. Each cell depicts the level of agreement between the grouped student answers.

which feedback messages are appropriate for groups of student responses. To help with this challenge, we have been working with teachers to construct datasets which contain multiple teachers looking at the same sets of student responses. Each teacher was asked to categorize student responses and write a feedback message that is appropriate and meaningful for all responses within the identified category; while not individualized, such a task brings us one step closer to providing directed feedback to students with similar answers (i.e. groups of students exhibiting the same misconception) in a similar manner in which feedback can be given to common wrong answers (c.f. Section 1.2). With multiple teachers performing this task, we are beginning to identify where teachers agree and disagree regarding how to define and group similar student responses; some teachers may define similar responses differently than others, and understanding this variance may not only help to construct a model to suggest responses, but also can help incorporate diversity among the suggestions.

Heatmaps are used to visualize the overlap of teacher categorization of answers for each problem. An example of such a heatmap is illustrated in Figure 5. Four teachers categorized the same student answers to a single open-ended problem. Dark cells indicate high agreement (there was high agreement as to which student answers would belong to the same group) while light cells indicate low agreement; for example, the “IDK” category of Teacher 4 has high agreement with the “NAN” category created by Teacher 2. The number of categories also highlight what each teacher felt were important distinguishing factors; Teacher 1, for example, felt it sufficient to only create two categories based on the specificity of student responses.

5 THE FUTURE OF GENERATING DIRECTED FEEDBACK

The development of the underlying machine learning models, as well as the software to support directed feedback messages through QUICK-Comments, is only half of the implementation cycle. There are three metrics that are planned to help evaluate the success of the tool. The first is that the tool should be, as the name implies, quick; teachers using the tool should be able to more quickly assess and provide feedback to their students. If successful, it is our hope that a larger percentage of assigned open-ended problems will be scored and feedback will be provided to more students through the use of the tool.

The second metric is an indicator of how well our model is able to suggest feedback messages. Teachers’ decisions to select or ignore our suggested messages provides a clear indicator as to how appropriate the messages are for the associated student answers. Our goal is that the tool will help provide more personal feedback to students, and that teachers will trust the tool to make appropriately helpful suggestions.

Finally, and we argue most importantly, the third metric is in regard to whether the suggested feedback messages are actually beneficial to students. Assessing whether students’ performance in the given content improves as a result of receiving feedback from their teacher provides an indicator as to whether the feedback is meaningful.

Toward these goals, as the tool is entering its pilot phase of testing, there are still many challenges to overcome in order to provide suggested feedback that is specific, personalized, and pedagogically sound. Continued collaboration with teachers in this process is a necessary aspect of developing and

improving the tool and highlights the importance of including teachers in the design of educational systems.

ACKNOWLEDGEMENTS

We acknowledge and thank Cindy Starks, Dr. Hilary Kreisberg, Cristina Heffernan, and the teachers who have collaborated with us to make this research possible. We thank multiple NSF grants (e.g., 1931523, 1940236, 1917713, 1903304, 1822830, 1759229, 1724889, 1636782, 1535428, 1440753, 1316736, 1252297, 1109483, & DRL-1031398), the US Department of Education Institute for Education Sciences (e.g., IES R305A170137, R305A170243, R305A180401, R305A120125, R305A180401, & R305C100024) and the Graduate Assistance in Areas of National Need program (e.g., P200A180088 & P200A150306), EIR, the Office of Naval Research (N00014-18-1-2768 and other from ONR), and Schmidt Futures.

REFERENCES

- Botelho, A. F., Baker, R. S., & Heffernan, N. T. (2017, June). Improving sensor-free affect detection using deep learning. In *International Conference on Artificial Intelligence in Education* (pp. 40-51). Springer, Cham.
- Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4(4), 253-278.
- Erickson, J. A., Botelho, A. F., McAteer, S., Varatharaj, A., & Heffernan, N. T. (2020, March). The Automated Grading of Student Open Responses in Mathematics. In *Proceedings of the 10th International Conference on Learning Analytics & Knowledge* (In Press).
- Kannan, A., Kurach, K., Ravi, S., Kaufmann, T., Tomkins, A., Miklos, B., ... & Ramavajjala, V. (2016, August). Smart reply: Automated response suggestion for email. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 955-964).
- Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., & McClosky, D. (2014, June). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* (pp. 55-60).
- Paquette, L., Baker, R. S., de Carvalho, A., & Ocumpaugh, J. (2015, June). Cross-system transfer of machine learned and knowledge engineered models of gaming the system. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 183-194). Springer, Cham.
- Pashler, H., McDaniell, M., Rohrer, D., & Bjork, R. (2008). Learning styles: Concepts and evidence. *Psychological science in the public interest*, 9(3), 105-119.
- Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).
- Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. In *Advances in neural information processing systems* (pp. 505-513).
- Ramos, J. (2003, December). Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning* (Vol. 242, pp. 133-142).

- Razzaq, L. M., & Heffernan, N. T. (2009, July). To Tutor or Not to Tutor: That is the Question. In *Dimitrova, Mizoguchi, du Boulay & Graesser (Eds.) Proceedings of the 2009 Artificial Intelligence in Education Conference*. (pp. 457-464).
- Roschelle, J., Feng, M., Murphy, R. F., & Mason, C. A. (2016). Online mathematics homework increases student achievement. *AERA Open*, 2(4), 2332858416673968.
- Selent, D., & Heffernan, N. (2014, June). Reducing student hint use by creating buggy messages from machine learned incorrect processes. In *International conference on intelligent tutoring systems* (pp. 674-675). Springer, Cham.
- Yin, B., Patikorn, T., Botelho, A. F., & Heffernan, N. T. (2017, April). Observing personalizations in learning: Identifying heterogeneous treatment effects using causal trees. In *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale* (pp. 299-302).

Opportunities for Human-AI Collaborative Tools to Advance Development of Motivation Analytics

Steven C. Dang

Carnegie Mellon University, Human Computer Interaction Institute
stevenda@cs.cmu.edu

Kenneth R. Koedinger

Carnegie Mellon University, Human Computer Interaction Institute
koedinger@cmu.edu

ABSTRACT: Modern educational technology products must support learners' cognitive and motivational needs. Extending models of learner motivation to new products leads to difficulties in generalizing existing models to drastically different systems or operationalizing existing behavioral theories with sufficient construct validity. Overcoming these challenges requires teams with both learning science and data science skillsets. Recent advances in automated machine learning and interpretable machine learning have led to opportunities to empower learning scientists with the capabilities of an interdisciplinary data science team. Through a review of prior studies on motivation analytic development, we identify common data science challenges and review some successful algorithmic solutions. We also identify challenges to scaffolding these tasks to users without data science backgrounds and highlight some advances in automated machine learning and interpretable machine learning that may enable development of tools and services to fill this need.

Keywords: Motivation, Self-Regulated Learning, Automated Machine Learning, Interpretable Machine Learning, Measurement, Online learning environments

1 INTRODUCTION

There are many challenges in developing motivational analytics for online learning environments. Developers must go beyond analytics of easily observable interactions and utilize models that take into account the affective and self-regulated learning dynamics of learners in order to draw inferences on their motivations (Eccles & Wigfield, 2002). Developing high quality analytics requires accounting for a range of concerns around construct validity and reliability while also leveraging complex modeling methods that lie outside the realm of theories of learner motivation (Milligan, 2018). An open challenge in the measurement of motivation, like self-regulated learning (SRL), lies in the challenge of how to operationalize constructs on different systems (Roll & Winne, 2015). Overcoming these design challenges requires a mixture of learning science knowledge and data science skills. This can be a significant barrier for many educational technology product companies that do not have the resources to hire experts with such specialized skillsets. With recent advances in technologies such as interpretable and automated machine learning, there is an opportunity to make data science skills accessible to learning scientists and dramatically increase the pool of individuals capable of developing high quality motivational analytics. In this paper, we review prior work in analytic development for measurement of motivational constructs using strictly log data. In this work, we identify particular challenges to developing measures of motivational constructs and identify specific areas of opportunity

for tools and algorithmic development that can greatly lower the barriers to development of motivational analytics.

2 CHALLENGES IN MOTIVATION MEASUREMENT

Learners' motivation is a product of their SRL, and as highlighted by Winne (2010), SRL is contextual and context evolves as learners regulate their learning. The challenge of measuring motivation requires inferring many latent contextual influences such as learners' goals, metacognition, and task value (Winne & Hadwin, 2008). Many of these factors can be difficult to measure, but learner's displayed learning behaviors and strategies can be indicators of their latent motivational factors as evidenced in work by Dang & Koedinger (2019a). Prior work in learning analytics and educational data mining has elaborated models of affect, SRL strategies, and relevant learning behaviors (Lang et al, 2017). However, extending these models to new systems and datasets introduces a host of new challenges (Winne, 2014).

For instance, Rowe et al (2009) extended a model of off-task behavior (Baker et al, 2004) to a narrative-centered learning environment. This open-ended environment involved actions such as navigating a character around a virtual world, interacting with objects, and talking with non-player characters. In lieu of a usable measurement model, the authors developed an alternate operationalization of off-task behavior that required insight into the pedagogical value of possible interactions in the game. Performing this task required a degree of learning science knowledge that is typically not found in many data scientists in the work force.

Rowe et al were motivated to develop an off-task behavior measurement model by the findings of Baker et al, indicating that student learning is negatively impacted by such behaviors. To test their hypotheses, the authors collected additional motivational survey and achievement data. Exploratory analysis demonstrated that students engaged in off-task behavior about 15% (SD=8.9%) of the time, which differs from the 20% frequency found by Baker et al. The statistical analysis indicated that unlike Baker's off-task model, the Rowe model was not significantly related to pre-post learning. Likewise, the results found no relationship between off-task behavior and either achievement orientation or self-efficacy, contradicting the results of Dang & Koedinger (2019a). This analysis demonstrates two common challenges in extending theory to new systems. Performing such a validation requires a set of data science skills that are lacking in many industry learning scientists.

Also, despite appearing to match the Baker et al construct on its face, the evidence indicates that the model developed by Rowe et al was not measuring the same construct. This construct validity problem challenge is an open problem in scaling the research of the learning analytics and educational data mining communities to more online learning environments (Huggins-Manley et al, 2019). One common process for validating a construct on a system is to collect ground truth data, as done by Rowe et al, and to leverage this information to validate a proposed model by applying appropriate statistical tests for agreement with expectations. As online learning environments expand their available content and are deployed to broader audiences, the challenge becomes how to sample a representative dataset to train sufficiently general models, an increasingly cost prohibitive task.

3 OPPORTUNITIES FOR HUMAN-AI COLLABORATION

3.1 Leveraging Multiple Facets to Bootstrap Construct Validity

One challenge in using observed behaviors to measure some latent motivational construct is that, unlike in experimental contexts, multiple constructs are likely implicated in any given behavior. Huggins-Manley et al (2019) discuss several relevant threats to construct validity. Construct confounding occurs when “inferences are drawn on one construct even though indicators reflect more than one construct”. Confounding constructs with facets of constructs occurs when “only some facets of a construct are measured, invalidating inferences about the full construct”. Mono-operation bias occurs because “a single indicator of a construct underrepresents the inferred construct, which is more complex than a single indicator”.

In our prior work, we attempted to tackle these threats to construct validity by leveraging multiple indicators of the target construct (Dang & Koedinger, 2019b). In this work, we operationalized our latent construct, diligence, through a series of metrics defined around how readily learners start work and how long learners can maintain focus. Analysis demonstrated that combined factors yielded both better predictions, reliability, and alignment with motivational factors as measured through correlation with survey-based motivation instruments. These results point to an interesting foundation for tools and services that support a construct operationalization process by leveraging multiple facets of a construct defined in the available behavior data in lieu of ground truth labels to perform construct validation and iteration.

3.2 Fitting Parameters using a Multi-faceted Latent

Operationalizing a measurement model for classifying a target behavior is a complex process involving a combination of expert learning science knowledge to understand the types of constructs to target in the data and data science knowledge to identify how to measure such targets in the data. Aleven et al (2006) demonstrate an a-priori thresholding process for operationalizing SRL theory into a measurement model on fine-grained data. The model consists of a set of if-then-else rules representing a decision tree model for a pattern of help seeking hypothesized by SRL theory. In order to apply this model to the data, the model operationalized concepts such as “Familiar at all?” and “Sense of what to do?” using a set of calculated values in the data and thresholds that are set to values that the authors describe as “intuitively plausible, given our past experience”. This a-priori heuristic is difficult to reproduce and requires an intuitive sense of how users interact with the system, which is not necessarily experience many product development teams possess.

Baker et al (2004) demonstrate a typical approach in the machine learning community, treating the model definition problem as a supervised machine learning problem. Ground truth labels of gaming the system behavior were collected from classroom observations and were used to train a machine learning model on the data. This model fitting process requires a degree of data science experience to both identify the correct algorithm to apply to the data given an understanding of the deeper structure of the problem and to elaborate a set of raw features from the raw data that can improve the ability the algorithm to find a well-fitting model. New toolkits in the automated machine learning (auto-ml) community have simplified this process of feature engineering. For instance, Kanter & Veeramachaneni

(2015) developed the Featuretools framework that leverages deep machine learning algorithms to automatically elaborate meaningful features over the raw data and takes a user-defined goal to automatically identify appropriate algorithms to fit a model that solves the target problem.

Another challenge in applying machine learning algorithms is evident in how Kuvalja et al (2014) leverage a pattern recognition algorithm to identify patterns of behavior that were indicative of SRL processes. In order to apply the algorithm, the authors defined three parameters: the minimum number of occurrences of a pattern, the probability of observing the pattern, and a threshold for how often a pattern must be observed in some time interval. Setting values for these algorithm parameters appears similar to the a-priori threshold setting demonstrated by Aleven et al (2006) but requires both knowledge relevant to the occurrence of SRL behaviors in practice and an understanding of the algorithm. Work in auto-ml also tackles this problem of automated parameter selection. For instance, Kandasamy et al (2019) leverage bayesian optimization to automatically identify the optimal parameter values to use for a particular machine learning algorithm to fit a model to a given data set.

Beyond the issue of lack of ground truth labels discussed in section 3.1, building high quality models of behavior for an online learning environment requires a number of other data science skills not typical of the training for many learning scientists. Advances in auto-ml have demonstrated a capacity for intelligent algorithms to tackle many of these tasks with minimal input from users, making such tasks more accessible to non-data science users. While a broader survey of auto-ml is beyond the scope of this work, we point to Zöller & Huber (2020) for a more comprehensive survey of available auto-ml frameworks and their capabilities.

3.3 Identifying Heterogeneity

Many online learning environments leverage fine-grained moment-by-moment behavior data to inform analytics. However, the contribution of Aleven et al (2006) highlights how many learning science theories do not make strong predictions about exactly how motivational factors influence learner's decision-making given some specific combination of contextual factors that we can observe in the data and bridging this gap is a not trivial. Data scientists aggregate learner behaviors to test theoretically predicted relationships that should be evident across contexts. Such aggregation techniques make it difficult to identify where a model may be inadequately capturing the target construct. For instance, a disengaged learner is expected to be lower performing than a more engaged learner because learning necessarily requires completion of some work to engage with concepts. Engagement might reasonably be operationalized as total time on task. Shih et al (2011) demonstrate that the quality of a student's engagement is evident in the speed of a student's response to a problem immediately following a request for help. Thus, two students might appear very similar in their total time working, but within that time, students' varying levels of cognitive engagement is evident in the differences in response time following instructional assistance provided by the learning environment. Using a simple operationalization of engagement would miss this source of variation in the data. Available model performance metrics and model interpretation tools lack adequate support for

learning scientists to think critically about the performance of their current models and identify these shortcomings (Kaur et al, 2019).

3.3.1 *Scaffolding Model Iteration with Qualitative Data Analysis*

We believe that data programming (Ratner et al, 2016) offers an interesting approach to allowing users to encode expert knowledge onto datasets. In this paradigm, users are asked to define rules, similar to the if-then-else rules defined by Alevi et al (2006), that encode some heuristic that experts might leverage to make classification judgements when reviewing learner behavior. These rules can be collected across multiple users and be overlapping and partially disagree with each other. The set of rules are used to infer labels for the data. While this is an interesting approach for enabling users to more naturally encode their knowledge onto the data, there remains the question of how to support users in realizing what knowledge might be relevant?

Baker & de Carvalho (2008) demonstrated that users with learning science knowledge and product familiarity are capable of reviewing segments of learner behavior data and drawing valid inferences on what that learner may be doing. This is a viable method to leverage learning scientists to identify instances of learner behavior that may be incorrectly classified and activating relevant knowledge that could be used to define heuristic rules to describe the model shortcoming. However, there are several barriers to supporting users in searching for segments of learning actions that may reflect some unknown but currently unaccounted for adjustment to the current model. We review some of the challenges we have faced in applying qualitative analysis to iterate on models of motivation and highlight opportunities for intelligent algorithms and tools to improve this workflow.

3.3.2 *Prioritizing Data for Qualitative Review*

As discussed previously, learning science theory typically can only make predictions about learner behavior when aggregated to student-level analytics. However, the goal of the qualitative analysis process is to support experts by reviewing cases of fine-grained behavior. This implies a two-step process where users must first identify a student to review and from that student's data, specific instances of behavior must be selected for review. Datasets can involve hundreds if not thousands of students, and just an hour of student data can translate to several hundred data points. There is an opportunity to improve the efficiency of this exploration process by leveraging algorithms to prioritize data to review.

In the first stage of the process, users need support in identifying which student's data to analyze first. Learning science theory informs users expectations for relationships between student-level aggregated variables in the data. These relationships can be used by anomaly detection algorithms to associate each student with some degree of non-fit to expectation and prioritize students accordingly. Anomaly detection can be as simple as defining a linear regression predicting some relationship between the available data and using the size of the prediction error as a data ranking. More complex anomaly detection methods are available, and we reference Chandola et al (2009) for a more comprehensive review of this literature. We believe tools that can provide such a ranking mechanism in addition to an ability to review metadata for each student would greatly improve users' ability to more effectively decide how to explore available student data.

Within the subset of data from an individual learner, the next step of the qualitative analysis process is to select which behavioral data to review first. The greatest challenge in this space is in supporting a search process where the target of the search is unknown. Anomaly detection algorithms can be applied to the raw data and then the data could be prioritized within student based on how anomalous the data appears to be. However, such methods do not leverage unencoded expert knowledge to support the ranking process. Recent SRL research has applied pattern-mining methods to identify relevant behaviors because SRL behaviors are driven by learning processes and which causes observable reoccurring patterns in the data (Molenaar & Järvelä, 2014). We believe it would be valuable to cluster and summarize students' behaviors and then to leverage interpretable machine learning frameworks such as interpret ML (Nori et al, 2019). These frameworks can enable users to understand the type of behavior encapsulated by a cluster in terms of the key contextual and behavioral features in the data. Similar to the method proposed by Baker & de Carvalho, expressing the data summaries in terms of these low-level details can allow experts to infer what may be happening and identify possible unexpected relationships. Together, these anomaly detection and interpretable machine learning techniques offer new avenues to surface relevant structure in the data that can inform learning scientists while searching through the vast quantity of learner data.

4 DISCUSSION

The major challenge in motivational measurement lies in identifying developers with both the data science and learning science knowledge to competently build measurement models from existing prior work in the learning analytics and educational data mining fields. We have demonstrated through examples in our prior work as well as others that there are opportunities for applying advances in autonomous machine learning and interpretable machine learning to empower learning science experts without experience in data science to be able to perform the same construct operationalization processes for building motivational analytics. We believe this is a promising open area of research for tools and algorithm development that can greatly accelerate the adoption of motivational analytics in online learning environments throughout the marketplace.

REFERENCES

- Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor. *International Journal of Artificial Intelligence in Education*, 16(2), 101-128.
- Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004, April). Off-task behavior in the cognitive tutor classroom: when students "game the system". In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 383-390).
- Baker, R., & de Carvalho, A. (2008, June). Labeling student behavior faster and more precisely with text replays. In *Educational Data Mining 2008*.
- Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in "gaming the system" behavior in interactive learning environments. *Journal of Interactive Learning Research*, 19(2), 185-224.

- Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3), 1-58.
- Dang, S., & Koedinger, K. R. (2019a). Exploring the Link Between Motivations and Gaming. In *Proceedings of The 12th International Conference on Educational Data Mining (EDM 2019)* (pp. 276-281).
- Dang, S., & Koedinger, K. R. (2019b). Towards a Behavior-Based Psychometric Instrument for Unobtrusive Measurement of Diligence. Manuscript submitted for publication.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual review of psychology*, 53(1), 109-132.
- Huggins-Manley, A. C., Beal, C. R., D'Mello, S. K., Leite, W. L., Cetin-Berber, D. D., Kim, D., & McNamara, D. S. (2019). A commentary on construct validity when using operational virtual learning environment data in effectiveness studies. *Journal of Research on Educational Effectiveness*, 12(4), 750-759.
- Kandasamy, K., Vysyaraju, K. R., Neiswanger, W., Paria, B., Collins, C. R., Schneider, J., ... & Xing, E. P. (2019). Tuning hyperparameters without grad students: Scalable and robust Bayesian optimisation with Dragonfly. *arXiv preprint arXiv:1903.06694*.
- Kanter, J. M., & Veeramachaneni, K. (2015, October). Deep feature synthesis: Towards automating data science endeavors. In *2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (pp. 1-10). IEEE.
- Kaur, H., Nori, H., Jenkins, S., Caruana, R., Wallach, H., & Vaughan, J. W. (2019). Interpreting Interpretability: Understanding Data Scientists' Use of Interpretability Tools for Machine Learning. Conditionally accepted at the 2020 ACM CHI Conference on Human Factors in Computing Systems (CHI 2020)
- Kuvalja, M., Verma, M., & Whitebread, D. (2014). Patterns of co-occurring non-verbal behaviour and self-directed speech; a comparison of three methodological approaches. *Metacognition and learning*, 9(2), 87-111.
- Lang, C., Siemens, G., Wise, A., & Gasevic, D. (Eds.). (2017). *Handbook of learning analytics*. SOLAR, Society for Learning Analytics and Research.
- Milligan, S. K. (2018, March). Methodological foundations for the measurement of learning in learning analytics. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 466-470).
- Molenaar, I., & Järvelä, S. (2014). Sequential and temporal characteristics of self and socially regulated learning. *Metacognition and Learning*, 9(2), 75-85.
- Nori, H., Jenkins, S., Koch, P., & Caruana, R. (2019). InterpretML: A Unified Framework for Machine Learning Interpretability. *arXiv preprint arXiv:1909.09223*.

- Paquette, L., de Carvalho, A. M., & Baker, R. S. (2014, July). Towards Understanding Expert Coding of Student Disengagement in Online Learning. In *CogSci*.
- Ratner, A. J., De Sa, C. M., Wu, S., Selsam, D., & Ré, C. (2016). Data programming: Creating large training sets, quickly. In *Advances in neural information processing systems* (pp. 3567-3575).
- Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, 2(1), 7-12.
- Rowe, J. P., McQuiggan, S. W., Robison, J. L., & Lester, J. C. (2009, July). Off-Task Behavior in Narrative-Centered Learning Environments. In *AIED* (pp. 99-106).
- Shih, B., Koedinger, K. R., & Scheines, R. (2011). A response time model for bottom-out hints as worked examples. *Handbook of educational data mining*, 201-212.
- Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational psychologist*, 45(4), 267-276.
- Winne, P. H. (2014). Issues in researching self-regulated learning as patterns of events. *Metacognition and Learning*, 9(2), 229-237.
- Winne, P. H., & Hadwin, A. F. (2012). The weave of motivation and self-regulated learning. In *Motivation and self-regulated learning* (pp. 309-326). Routledge.
- Zöller, M. A., & Huber, M. F. (2019). Survey on automated machine learning. *arXiv preprint arXiv:1904.12054*.

Sphinx: An Automated Generation System for English Reading Comprehension Assessment

Saad Khan (saad.khan@act.org) ,Yuchi Huang (yuchi.huang@act.org), Scott Pu, Vladimir Tarasov, Alejandro Andrade, Richard Meisner, Dave Edwards, Alina von Davier

ABSTRACT: We discuss Sphinx, a human-AI hybrid system for scalable production of reading comprehension passages in English from writers' samples/prompts to be used in a variety of learning and assessment. To the best of our knowledge, Sphinx is the first natural language generation system designed to create reading passages in a computationally efficient manner and can be used in a plethora of learning and assessment contexts. In Sphinx, we integrate state-of-the-art NLP approaches with the reasoning ability of writers to process text from a multiple of sources and produce industry-grade quality narratives and original content at the same time. We utilize highly capable NLP transformer models such as BERT, GPT2 and USE to encode text data and automate writer's tasks including, but not limited to topic modeling, auto-summarization, sentence recommendation and ranking, and paraphrasing. Furthermore, we integrate AEGIS (ACT English Item Generation System) into Sphinx to repeatedly produce items from source and composed essay text. Along with questions of quality and rigor, we pay special attention to the issues of parallelization and scalability that are pressing in the learning and assessment industries.

Format: Practitioner Presentation

Keywords: Natural Language Processing, Automated Content Generation

1 INTRODUCTION

Educational assessment, learning, and publishing companies dedicate significant resources for the creation of original text-based content for use in formative and summative tests, as well as in classroom learning or open educational resources. This process can be laborious, highly dependent on domain expertise and difficult to scale up. Furthermore, the manual generation of content and assessment items heightens the risk of incomplete, duplicate and/or redundant content. Automating educational content generation such as assessment items and in particular English reading passages (see figure 1 for a sample) can result in cost savings, quality standardization, and open new possibilities for personalized learning experiences.

Classical natural language processing (NLP) work in this area dates back to John Wolfe's seminal work (Wolfe, 1977) that demonstrated the feasibility of automatically generating natural language questions. In recent years there has been a revival in interest, spurred in part by advances in dialogue systems such as Amazon Alexa. While

Questions 8-9 are based on the following passage.

In the summer of 1911, the explorer Hiram Bingham II bushwhacked his way to a high ridge in the Andes of Peru and beheld a dreamscape out of the past. There, set against

Line looming peaks cloaked in snow and wreathed in clouds,

5 was Machu Picchu, the famous "lost city" of the Incas. This expression, popularized by Bingham, served as a magical elixir for rundown imaginations. The words evoked the romanticism of exploration and archeology at the time. But finding Machu Picchu was easier than

10 solving the mystery of its place in the rich and powerful Inca empire. The imposing architecture attested to the skill and audacity of the Incas. But who had lived at this isolated site and for what purpose?

8. The words "magical elixir" (line 7) primarily emphasize the

- (A) motivation for an expedition
- (B) captivating power of a phrase
- (C) inspiration behind a discovery
- (D) creative dimension of archaeology
- (E) complexity of an expression

Figure 1: Sample reading passage and associated item. Manual creation of such passages can be a costly and inefficient process.

traditional approaches to NLP-based educational item generation involve a pipeline of modules such as content selection, template design and item realization (Gierl et al., 2012;), these have been criticized for being rigid and too reliant on arbitrary heuristic rules (Heilman, 2011). There is growing interest in developing end-to-end deep neural network based approaches that do not require customized, hand crafted rules and are better

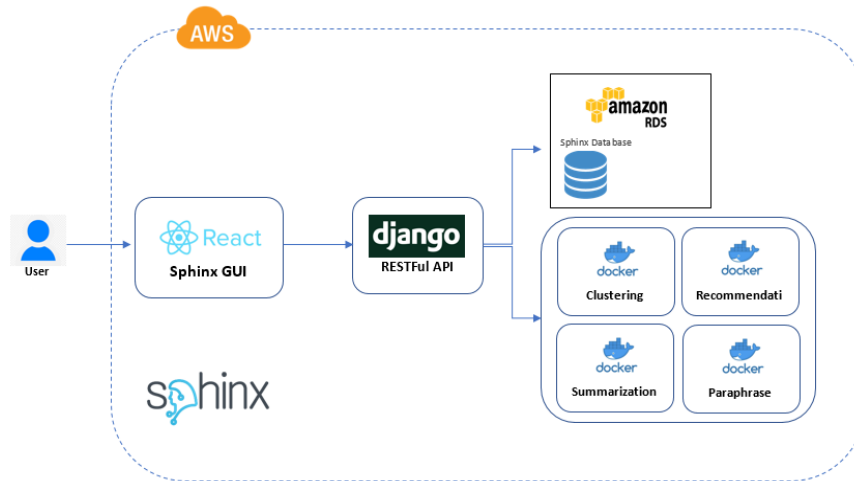


Figure 2: Sphinx system architecture is designed to be modular with distributed services hosted on AWS..

equipped to generalize across content areas (Cervone et al. 2019). A key element of such approaches is leveraging large text content databases and well annotated datasets such as BookCorpus (Zhu et al. 2015), SQuAD (Rajpurkar et al., 2016) and Wikipedia.

In this paper we discuss Sphinx, a scalable system that utilizes advanced NLP models to help expert or novice writers interactively create English reading comprehension passages from writers' samples/prompts. To the best of our knowledge, Sphinx is the first natural language generation system designed to create reading comprehension passages in a computationally efficient manner. Passages created by Sphinx could be used in a variety of learning and assessment applications such as formative and summative assessments of reading and comprehension and real-time adaptive learning. The system is designed to integrate state-of-the-art NLP approaches with the reasoning ability of writers to process text from a multiple of sources and produce industry-grade quality narratives and original content at the same time. Highly-capable NLP transformer models such as BERT (), GPT2 () and USE () are utilized to encode text data and automate writer's tasks including, but not limited to topic modeling, auto-summarization, sentence recommendation and ranking, and paraphrasing. The recommendations of passage content made by NLP models will always be evaluated by human users before inclusion. This interactive feature enables quality control for improved content validity as well as collecting training data for underlying machine learning models. We believe such human-AI hybrid systems can be the best of both worlds by utilizing the reasoning ability of subject matter experts while processing large amounts of input text to automate portions of the writing process. Furthermore, we integrate AEGIS (ACT English Item Generation System) into Sphinx to repeatedly produce items from source and composed essay text. Along with questions of quality and rigor, we pay special attention to the issues of parallelization and scalability that are pressing in the learning and assessment industries.

In the following, we describe the system architecture and the technology stack used in development. Especially, we present detailed framework of the Sphinx system, including its core NLP algorithmic modules: automated summarization, topic clustering, sentence recommendation and paraphrasing. We also introduce AEGIS (ACT English Item Generation System) at the end of the second section, and followed by the conclusion and discussion on the future work finally.

2 TECHNICAL MODULES

Figure 2 shows an overview of the Sphinx system architecture. Sphinx is designed as a distributed system with three main components. The first is a React JavaScript-based graphic user interface (Figure 3). Through the interface, users can upload or download passages, organize passages into folders, create projects, and use Sphinx's NLP sub-modules to compose new passages. Users can also choose whether to enable the scaffold feature, so that new article composition will be divided into three parts: Introduction, Body and Ending. The user interface is linked to the second component, which is a Django REST API (<https://www.django-rest-framework.org/>). The API serves mainly as a gateway to the NLP machine learning algorithms. It authenticates user visits, manages processing requests and access to the system database. Sphinx's NLP algorithms form the third core component and include text summarization, topic modeling, sentence recommendation and paraphrasing. As shown in Figure 4, for most functionality modules we provide different NLP algorithms for users to choose from. Each algorithm is wrapped as a web service in a docker container (<https://www.docker.com/resources/what-container>) and remains dormant unless a processing request is received through the Django REST API.

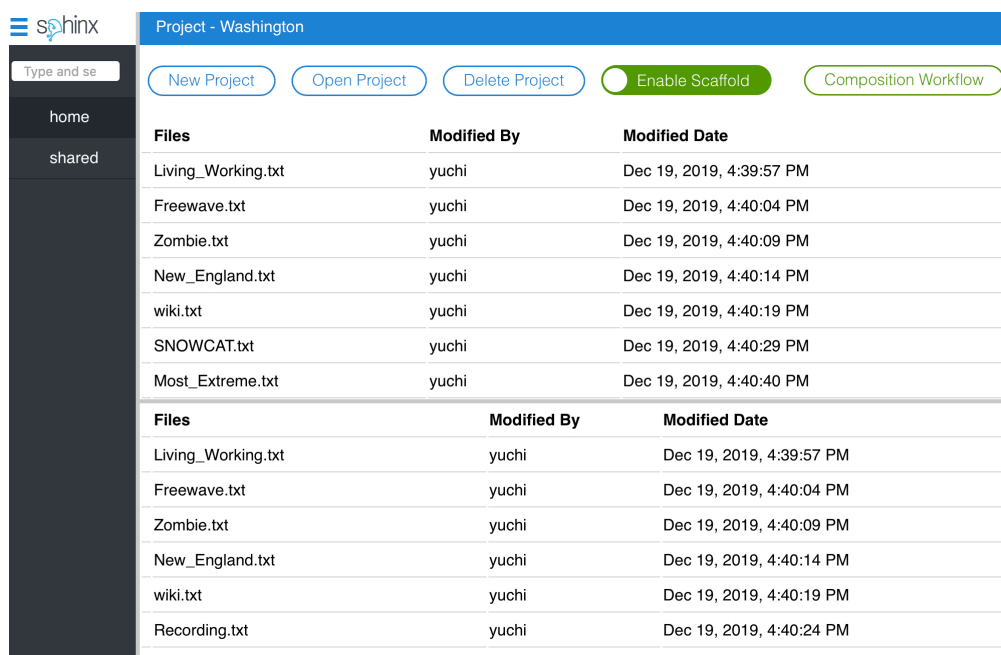


Figure 3: Through the interface, users can upload or download passages, organize passages into folders, create projects, and use Sphinx's NLP sub-modules to compose new passages.

Currently, Sphinx's server and algorithms are all deployed on AWS EC2 instances. AWS provides auto-scaling services that can add or remove EC2 instances dynamically according to real (or predicted)

demand. AWS also comes with services like Elastic Loading Balancing that automatically distributed income traffic to different EC2 instances according to their current workload. Technologies like Elastic Beanstalk can take care of auto-scaling, load balancing, application health monitoring, and more according to configuration.

Engine	Algorithm	Options (range of values)
Sentence Recommendation Model	<input checked="" type="radio"/> graph <input type="radio"/> bert	5 Number of sentences
Summarization Model	<input checked="" type="radio"/> bertsum <input type="radio"/> textrank	0.5 Ratio (0-1)
Topic Extraction Model	<input checked="" type="radio"/> topicrank <input type="radio"/> yake <input type="radio"/> rake	3 Number of keywords 4 Max words
Clustering	<input checked="" type="radio"/> kmeans	5 Number of Topic Clusters
Paraphrase	<input checked="" type="radio"/> backtraslation	3 Number of paraphrases <input checked="" type="checkbox"/> Ner

Figure 4: For most functionality modules we provide different NLP algorithms for users to choose from.

In the following we describe details of the NLP algorithmic core of Sphinx, functionalities that enable expert or novice writers to process raw digital text from a multitude of sources into new, coherent narratives and original reading content. We also introduce AEGIS (ACT English Item Generation System) which has been integrated to Sphinx as an essay-based item generation tool.

2.1 Automated Summarization

Living_Working.txt ✕

Summary	Content
<p>An infamous place of extremes, 6,288 foot Mount Washington, New Hampshire has been captivating audiences for hundreds of years.,Mount Washington Observatory has been operating a continuously-staffed scientific outpost on this remote peak since 1932, providing the Observatory many remarkable stories and an intimate knowledge of the mountain.,They have amassed one of North America's longest continuous climate records, and developed an intimate understanding of the place known as the "Home of the World's Worst Weather.",at the summit , participants are treated to a tour of the famous mountaintop weather station , an in-depth look at the work of the observatory , a home-style lunch mountain prepared by observatory volunteers , and time outdoors to experience the awe and wonder of mount washington 's brutal weather .,on</p>	<p>An infamous place of extremes, 6,288 foot Mount Washington, New Hampshire has been captivating audiences for hundreds of years. While many places on Earth experience severe weather, few are inhabited by humans 24 hours a day, 365 days a year. Mount Washington Observatory has been operating a continuously-staffed scientific outpost on this remote peak since 1932, providing the Observatory many remarkable stories and an intimate knowledge of the mountain. Through the years, the hardy men and women of the Observatory have experienced white-out blizzards, stunning 110-mile vistas, and everything in between. They have survived a 231mph wind gust, endless days of disorienting fog, and snow drifts more than 20 feet tall. They have worked with scientists and celebrities, students and snow rangers. They have amassed one of North</p>

Figure 5. Generated extractive summary (left) and a source document (right) shown in Sphinx interface.

A key feature of Sphinx is automated text summarization (Figure 5). At the outset users can upload a variety of original articles that can then be readily summarized into prototype passages for faster understanding and even use as draft text for new compositions. Text Summarization is an area of Natural Language Processing (NLP) which is bound to have a huge impact on a lot of applications such as media monitoring, newsletters, social media marketing among others. In this project, we focus on

extractive method in which a shorter paragraph is created by extracting and concatenating a subset of spans (usually sentences) from a document, so that the summarized information is as close to the original text as possible. Let a denote an article containing several sentences $[s_1, s_2, \dots, s_m]$, where s_i is the i th sentence in the document. Our problem is defined as the task of assigning a label $y_i \in \{0, 1\}$ to each s_i , indicating whether the sentence should be included in the summary. In our system, we adopted two extractive approaches. The first one is the most important early work and baseline for extractive summarization, named 'TextRank' (Mihalcea and Tarau, 2004), in which a graph-based ranking model similar to Google's PageRank (Brin and Page, 1998) was proposed to extract core sentences from text in real time. The second approach, BERTSum (Liu, 2019), is a state-of-the-art method which outperforms previous work on the CNN/Dailymail dataset (Hermann et al., 2015). BERTSum is fine-tuned on top of the famous BERT (Bidirectional Encoder Representations from Transformers) pretrained model (Devlin et al. 2019), which have recently advanced a wide range of natural language processing tasks. To apply BERT on text summarization, BertSum made the following changes to Bert: 1) Encoding multiple sentences in the input level by inserting a [CLS] token before each sentence and a [SEP] token after each sentence; 2) using interval segment embeddings EA (if i is odd) or EB (if i is even) to distinguish multiple sentences within an article; 3) On the sentence representation vectors T_i (the vector of the i th [CLS] symbol) of the top BERT layer, adding a linear classifier and using a sigmoid function to get the predicted score. After fine-tuning the pretrained BERT model on CNN/DailyMail news dataset, the BERTSum system is able to create high-quality extractive summarization for input articles.

2.2 Topic Modeling

The screenshot displays the Sphinx web application interface. On the left is a dark sidebar with a menu containing 'Project', 'Generated Passage', 'Introduction', 'Body', and 'End'. The main content area has a blue header 'Introduction Topic Clusters'. Below this header are three topic cards: 'Topic 1' (grey) with keywords 'mountain', 'washington', 'snow'; 'Topic 2' (grey) with keywords 'miles', 'hour', 'mount washington'; and 'Topic 3' (blue) with keywords 'mount washington observatory', 'weather', 'summits forecast'. Below the topic cards is a section titled 'Recommended Sentences' with tabs for 'Topic Seed', 'Source Archive', and 'NLG'. The 'Topic Seed' tab is active, showing a list of sentences with a '+' icon to the right of each. The sentences are: 'The observatory also provides a higher-summits forecast that applies to all mountains above 4,000 feet in New Hampshire, so that skiers, hikers and campers can plan their excursions while accounting for the forecast.', 'In addition to its meteorological and research endeavors, the Observatory is involved in many educational efforts which seek to inform individuals about the many significant aspects of weather, area history, and the mountain environment.', 'At the Mount Washington Observatory, staff measure current conditions every hour, including temperature, wind direction and speed, precipitation, cloud height and coverage, and more.', 'The weather observation station is located on the summit of Mount Washington in New Hampshire.', and 'The Observatory's mission is to advance understanding of the natural systems that create the Earth's weather and climate, by maintaining its mountain top weather station, conducting research and educational programs and interpreting the heritage of the Mount Washington region.' A 'Show More' button is at the bottom of the list.

Figure 6. Sphinx processes all sentences of raw material documents to generate topic clusters with their keywords (above) and corresponding core sentences (shown as 'Topic Seed' below).

In addition to creating summaries, Sphinx analyzes the raw material articles/documents input by the user to automatically identify and extract latent topics and related core sentences. Since new articles in Sphinx

are composed at sentence level, we perform clustering on sentences of raw articles and infer topic phrases on formed clusters. These topics serve as a guidance for writers to ensure comprehensive content coverage and the starting point for new composition. Our topic modeling approach begins with utilizing Google's Universal Sentence Encoder (Cer et al. 2018) to encode the sentences of all input articles into a 512-dimensional vector. This transformer encoder was trained from a large corpus composed from a variety of data sources with the aim of accommodating a wide variety of natural language understanding tasks such as text classification, sentimental analysis etc. Second, K-means clustering is conducted only on the embedded vectors of those summarized sentences of all articles. To avoid the influence of the trivial description in the original articles on the topic modeling, we only perform the clustering on extractive sentences generated from the text summarization step. Third, the encoded vectors of all sentences are projected into computed clusters and only top sentences closest to each cluster center are kept, ranked and presented to users. At last, three unsupervised topic extraction methods are integrated into our system for users to choose from: graph-based method TopicRank (Bougouin et al. 2013), YAKE (Campos et al, 2020) and RAKE (Rose et al. 2010). Each of three approaches can be employed on top sentences of a cluster to extract key phrases which cover the major topics depicted in those clustered sentences. If the scaffold mode is enabled, the topic sentences are reranked so sentences from specific part (e.g. the introduction) of raw documents are presented to users at a higher priority.

2.3 Sentence Recommendation

Once the user selects a topic cluster, Sphinx recommends a list of sentences of that cluster (the quantity is user configurable) from which the user can choose a seed sentence, as shown in Figure 6. In article composition, we adopt an interactive and recursive strategy in Sphinx to integrate the writing skills of human users and language processing abilities of machine learning algorithms. For the second sentence of a new composition, besides selecting a new sentence from the topic clusters, users can also choose from sentences recommended by Sphinx to fit the content (Figure 7). The recommended sentences could come from archived sentences of source documents (shown as 'Source Archive' in this figure) or new sentences generated from the GPT2 (Radford et al., 2019) model (shown as 'NLG' in this figure). This composition process then repeats until the user is satisfied with the content of the passage draft. Sphinx provides three recommendation engines, of which the first method has lower time complexity, the second method returns higher recommendation accuracy from achieved source sentences, while the third method creates new sentences not from source documents. All three engines evaluate contiguity and cohesion to filter down the next sentence recommendation list to a user configured number, which ensures semantic meaning carries over in sentence transition. In the first engine, embedded vectors of all sentences are computed from Google's Universal Sentence Encoder and an affinity graph is created using cosine similarity measures. Based on the undirected manifold graph ranking algorithm (Zhou, 2004), the sentence recommendation problem is formulated as a sentence ranking problem given the query sentence (the last sentence in the composed new article). The second recommendation engine is based on a pretrained BERT (Devlin et al., 2019) model with the next sentence prediction objective, which was trained on pairs of sentences from a variety of datasets to specifically model the relationship of two input sentences - whether they are next to each other. In a transfer learning setting, we get rid of the output layer of binarized next sentence classification and utilize the last hidden layer features of tokens of two sentences in a pair: (S_i and S_{i+1} where $i = 1, 2, 3 \dots p$). In this way, the problem of sentence recommendation becomes ranking the

similarities between the BERT features of S_e and S_i . Instead of simply averaging on all token features before computing the cosine similarity of S_e and S_i , we calculate a weighted sum of token features for each sentence according to the frequency of a word: less frequent words contribute more to sentence feature than frequent words (e.g., is and do) and more unique relatedness could be captured between two sentences. In the third recommendation engine, the famous language generation model GPT2 (Radford et al., 2019) is applied to generate brand new sentences different from archived source articles. The authenticity and language quality of GPT2 generated sentences will be evaluated by content specialists before adoption.

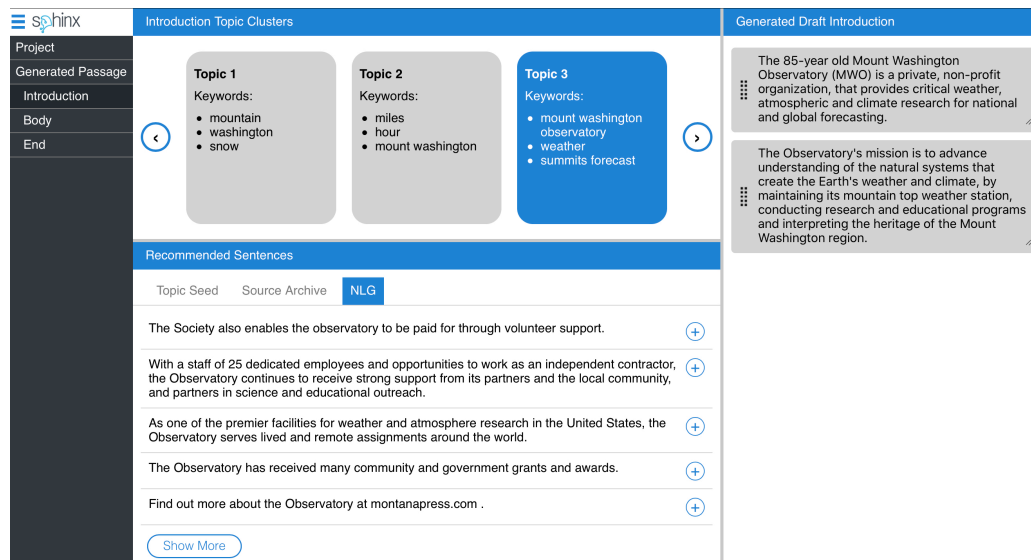


Figure 7. The recommended sentences could come from archived sentences of source documents (shown as ‘Source Archive’ in this figure) or new sentences generated from the GPT2 model (shown as ‘NLG’ in this figure).

2.4 Paraphrasing

At any given moment in draft composition, users have the option to use automated paraphrasing. Our sentence paraphrasing approach consists of three steps. First, a Named Entity Recognition and Discourse elements (NERD) filter masks all segments of the sentence that need to be kept intact, such as quotations and names of entities. Second, we use a back-translation approach whereby the sentence is translated to n different languages and then translated back to English using Google Translate (<https://translate.google.com/>) to generate paraphrase candidates. Third, each of the n sentence candidates are scored for their semantic similarity (USE cosine distance) and lexical-grammatical distance (Rouge Score, Conroy et al. 2006) with respect to the original sentence. A weighted average of these scores is used to rank list the paraphrased sentences and the user has agency to make a final selection. After all composed sentences are paraphrased, the user can continue to edit essays they work on before saving them (Figure 9).

Paraphrase ✕

In addition to its meteorological and research endeavors, the Observatory is involved in many educational efforts which seek to inform individuals about the many significant aspects of weather, area history, and the mountain environment.

- ☒ In addition to meteorological and research work, the observatory is also involved in many educational efforts in an effort to educate individuals about many important aspects of weather, regional history, and mountain environments.
- ☐ In addition to its meteorological and research efforts, the observatory participates in many educational efforts aimed at informing people about the many important aspects of the weather, the history of the area and the mountain environment.
- ☐ In addition to its meteorological and research efforts, the observatory is involved in a number of educational efforts to familiarize individuals with many important aspects of the weather, history of the area and the mountain environment.

Select

Figure 8: Sphinx’s paraphrasing module based on back-translation.

Composition Workflow

Essay Title Edit Essay Title
Save Draft Essay

Introduction

The 85-year old Mount Washington Observatory (MWO) is a private, non-profit organization, that provides critical weather, atmospheric and climate research for national and global forecasting. The observatory’s mission is to improve its understanding of the natural systems that make up the earth’s weather and climate by maintaining its peak weather stations, conducting research and education programs, and interpreting the heritage of the mount washington area. In addition to meteorological and research work, the observatory is also involved in many educational efforts in an effort to educate individuals about many important aspects of weather, regional history, and mountain environments.

Body

The only months of the year Mountain Washington has not recorded a below-zero temperature are June, July, August and September. The warmest temperature ever recorded at the summit of Mountain Washington is 72 degrees, which occurred in both August 1975 and June 2003. Even in the summer, July averages a trace of snow and August averages 0.1 inches of snow. In fact, 1.1 inches of snow was measured atop Mountain Washington in July 1957, and 2.5 inches was measured in August 1965. Known around the world for its incredible weather, winter on Mount Washington means freezing fog, heavy snow, bitter cold and winds that top the 100 mph mark on a regular basis. Mountain Washington’s snowiest winter was 1968-69 when 566.4 inches fell, including the 172.8 inches in February 1969, as referenced above. The average snowiest month at Mountain Washington is December, with an average of 45.5 inches of snow.

End

Part of the Observatory’s role involves gathering daily weather information. For over 13 years, Mount Washington Observatory has been using equipment from FreeWave, a specialist in machine-to-machine communications in industrial IoT (IIoT) environments, to help with weather monitoring across five remote weather stations. These stations and sensors measure temperature, humidity, wind speed and direction, and ground temperature, and require continuous connectivity to provide real-time weather feeds. Weather data is sent to the National Weather Service and becomes the basis for weather models used by all forecasters in our region.

Figure 9: An example of a draft composition. Users can continue to edit essays they work on before saving them.

2.5 Automated Item Generation

Rapidly evolving new directions in computer-adaptive tests with increased numbers of forms, and expanded markets, require a significantly higher level of item production. Automated item generation is a promising avenue for facilitating item development, though it has traditionally been limited to math content. Initial attempt of automatic English reading item generation at ACT focused on discrete item generation, which relied on content staffs to produce item text substrings that would continue to make sense in the context of the newly generated items. Recently, AEGIS (ACT English Item Generation System) has been created and successfully applied in the development of various ACT English Tests. In AEGIS, the possible components (e.g., the nouns, names, verbs, adjectives) of the generated items are not manually derived by content experts but are scanned for in the essay and transformed in rule-based ways to generate the item’s newly inserted essay linguistic error and item distractors. For example, Figure shows an essay excerpt and corresponding item with the item model classification SST-FOR-FRG, “Correcting rhetorically ineffective sentence fragments”:

Architect Eero Saarinen, who created the design that symbolized the memorial's theme of St. Louis as the "Gateway to the West."

- A. NO CHANGE
- B. Saarinen, creator of
- C. Saarinen created*
- D. Saarinen creating

Figure 10: An example of essay excerpt item which is classified under the item model SST-FOR-FRG, "Correcting rhetorically ineffective sentence fragments".

The comma and the "who" do not belong in the underlined portion of the essay, so the key is "C". As of the end of the year 2019, over 200 item models has been developed and put in operational usage; over 1000 items have been produced for a number of ACT English tests.

Since the item model is always based on the linguistic patterns, errors, and rules characterizing the abstract structure of a parent item of known high quality, the need for content expert involvement is minimal – the software automatically handles all of the actual item generation. For the same reason, AEGIS is particularly suitable to be integrated into large-scale content generation systems like Sphinx. As shown in Figure, AEGIS can be applied on both the source documents input by users or the newly-composed articles to generate and publish supported item types in real time, as shown in Figure 11.

The screenshot shows the Sphinx AEGIS interface. On the left, a source document is displayed with a paragraph about the Mount Washington Observatory (MWO). The text includes several underlined phrases: "weather, area history, and the mountain environment", "precipitation cloud height and coverage, and more", and "the highest wind speed, recorded in 1934, was 231 miles per hour".

On the right, the "Item models" section shows a table of generated items. The table has columns for PUN, USG, SST, KLA, and ORG. The items are listed in rows, with some highlighted in red and others in green. The items are:

PUN	USG	SST	KLA	ORG
AMB-001	RNR-007	UNP-012		
AMB-002	RNR-008	UNP-013		
AMB-003	RNR-009	UNP-014		
AMB-004	RNR-010	UNP-015		
ITS-001	RNR-011	WIS-001		
ITS-002	RNR-012	WIS-002		
ITS-003	RNR-013	WIS-003		
ITS-004	RNR-014	WIS-004		
POS-001	UNP-001	WIS-005		
POS-002	UNP-002	WIS-006		
POS-003	UNP-003	WIS-007		
POS-004	UNP-004	WIS-008		
POS-005	UNP-005	WIS-009		
RNR-001	UNP-006			
RNR-002	UNP-007			
RNR-003	UNP-008			
RNR-004	UNP-009			
RNR-005	UNP-010			
RNR-006	UNP-011			

Below the table, the "Parent item info" section shows the following details:

- Test: Aspire Grade 7
- Form: E702C1L0-SFT17
- IBN: Asp2016-WCONE-1624
- P-value: 73
- Biserial: .56
- DOK: 1

Figure 11: In AEGIS integrated into Sphinx, two items are generated under the item model 'ITS-004'.

3 CONCLUSION

In this paper we present Sphinx, a system that in a scalable and efficient manner helps writers compose English reading comprehension passages and corresponding items for use in educational

learning and assessment. We adopt an interactive and recursive strategy to integrate writing skills of human users and advanced NLP modules deployed in an auto-scaled manner. One of the benefits of our approach is that the human in the loop enables the AI to learn from expert writers as it accumulates useful data for future training of models, so make the large-scale learning and assessment more efficient, more effective and more adaptive. We believe the system can be used in delivering adaptive learning experiences as well as formative and summative assessments of English reading among other educational applications. Sphinx is currently in pilot trials and validity testing and we plan to introduce more advanced functionalities such as intelligent source document searching and organizing, structured article composition and automated figure generation among others.

REFERENCES

- Campos, R., Mangaravite, V., Pasquali, A., Jatowt, A., Jorge, A., Nunes, C. and Jatowt, A. (2020). *YAKE!:* *Keyword Extraction from Single Documents using Multiple Local Features*. In *Information Sciences Journal*. Elsevier, Vol 509, pp 257-289
- Cer, D., Yang, Y., Kong, S. Y., Hua, N., Limtiaco, N., John, R. S., & Sung, Y. H. (2018). *Universal sentence encoder*. arXiv preprint arXiv:1803.11175.
- Chen, D., Bolton, J., & Manning, C. D. (2016). *A thorough examination of the cnn/daily mail reading comprehension task*. arXiv preprint arXiv:1606.02858.
- Conroy, J. M., Schlesinger, J. D., & O'Leary, D. P. (2006, July). *Topic-focused multi-document summarization using an approximate oracle score*. In *Proceedings of the COLING/ACL*.
- Cervone, A., Khatri, C., Goel, R., Hedayatnia, B., Venkatesh, A., Hakkani-Tur, D., & Gabriel, R. (2019). *Natural Language Generation at Scale*. arXiv preprint arXiv:1903.08097.
- Gierl, M. J., Lai, H., & Turner, S. (2012). *Using automatic item generation to create multiple choice items for assessments in medical education*. *Medical Education*, 46, 757-765.
- Heilman, M. 2011. *Automatic factual question generation from text*. Ph.D. thesis, Carnegie Mellon University.
- Liu, Y. (2019). *Fine-tune BERT for Extractive Summarization*. arXiv preprint arXiv:1903.10318.
- Mihalcea, R., & Tarau, P. (2004). *Textrank: Bringing order into text*. In *Proceedings of the 2004 conference on empirical methods in natural language processing* (pp. 404-411).
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking*. Stanford infolab
- Radford, Alec and Wu, Jeff and Child, Rewon and Luan, David and Amodei, Dario and Sutskever, Ilya. *Language Models are Unsupervised Multitask Learners*. Technical report, OpenAi, 2019
- Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). *Squad: 100,000+ questions for machine comprehension of text*. arXiv preprint arXiv:1606.05250.
- Rose, S., Engel, D., Cramer, N., & Cowley, W. (2010). *Automatic Keyword Extraction from Individual Documents*. In M. W. Berry & J. Kogan (Eds.), *Text Mining: Theory and Applications*: John Wiley & Sons.
- Serban, I. V., García-Durán, A., Gulcehre, C., Ahn, S., Chandar, S., Courville, A., & Bengio, Y. (2016). *Generating factoid questions with recurrent neural networks*: arXiv:1603.06807.
- Wolfe, J.H. 1977. *Reading retention as a function of method for generating interspersed questions*. DTIC, 1977

- Zhu, Y., Kiros, R., Zemel, R., Salakhutdinov, R., Urtasun, R., Torralba, A., & Fidler, S. (2015). *Aligning books and movies*. In Proceedings of the IEEE ICCV (pp. 19-27).
- Zhou, D., Weston J., Gretton A., Bousquet O., and Schölkopf, B. *Ranking on Data Manifolds*. NIPS 2004.

Adaptive Learning Meets Crowdsourcing: Towards Development of Cost-Effective Adaptive Educational Systems

Hassan Khosravi

The University of Queensland

h.khosravi@uq.edu.au

ABSTRACT: A growing body of evidence demonstrates that adaptive educational systems (AESs) can provide an efficient, effective, and customised learning experience for students. Despite their success in enhancing learning, AESs may encounter barriers to adoption as they are generally expensive to develop, challenging to scale across disciplines, and face limitation in their ability to engage students in higher-order thinking. The use of crowdsourcing to support learning at scale and personalisation has recently received significant attention in the Artificial Intelligence in Education (AIED) and Educational Data Mining (EDM) communities. Building on this momentum, this short article considers the viability of using crowdsourcing as a way of addressing the abovementioned common AES challenges. We first discuss the viability and benefits of using crowdsourcing in adaptive educational systems. We then present a system called RiPPLE to demonstrate one approach for implementing a discipline-agnostic, cost-effective crowdsourced adaptive educational system that holds potential for promoting higher-order learning. We share initial results and lessons learned from piloting RiPPLE in 20 courses from 8 different disciplines and conclude by offering general implications and challenges for employing crowdsourcing within AESs.

Keywords: Adaptive educational systems, learner sourcing, crowdsourcing

INTRODUCTION

An adaptive educational system (AES) uses data about students, learning processes, and learning products to provide an efficient, effective, and customised learning experience for students. The system achieves this by dynamically adapting instruction, learning content, and activities to suit students' individual abilities or preferences (Aleven, McLaughlin, Glenn & Koedinger, 2016).

A consistent and growing body of knowledge over the past three decades has provided evidence about the effectiveness of AESs relative to traditional educational systems that offer instructions and learning activities that are not adaptive (Anderson, Boyle & Reiser, 1985; VanLehn, 2011; Ma, Adescope, Nesbit & Liu, 2014). Despite their ability to enhance learning, however, AESs have been embraced slowly by higher education, with adoption restricted mostly to research projects (Aleven et al., 2016; Essa, 2016).

To effectively adapt to the learning needs of individual students, an AES requires access to a large repository of learning resources. These resources are commonly created by domain experts. The development time for earlier versions of AESs is estimated at more than 50 hours of an expert's time for each hour of instruction (Aleven, McLaren, Sewall & Koedinger, 2006). Smart tools for authoring an AES, such as Cognitive Tutor Authoring Tools (Aleven et al., 2006; Aleven et al., 2016), have reduced the development time to roughly 25 hours of a domain expert's time per instructional hour.

Nevertheless, an AES is still very expensive to develop and challenging to scale across different domains.

How can institutions provide cost-effective AESs across many domains? One potential solution is to adopt a crowdsourcing approach, engaging students in the creation, moderation, and evaluation of learning resources (Heffernan et al., 2016). A crowdsourcing approach can significantly reduce development costs and has the potential to foster higher-order learning for students across many domains. But is this vision theoretically viable? Can students create high-quality resources? Are students able to effectively evaluate the quality of their peer-created resources? How does creating and evaluating resources impact learning? Following is an attempt to answer these critical questions.

CREATING AND MODERATING LEARNING RESOURCES IN PARTNERSHIP WITH STUDENTS

There seems to be adequate evidence suggesting that students can create high-quality learning resources that meet rigorous qualitative and statistical criteria (Walsh, Harris, Denny, Smith, 2018; Tacket et al., 2018; Galloway & Burns 2015; Bates, Galloway, Riise & Homer, 2014; Denny, Hamer & Luxton-Reilly, 2009). In fact, resources developed by students may have a lower chance of suffering from an expert blind spot (Nathan, Koedinger & Alibali, 2001). However, it seems likely that some learning resources developed by students may be ineffective, inappropriate, or incorrect (Bates, et al., 2014). Therefore, in order to effectively utilise resources developed by students, a selection and moderation process is needed to ensure the quality of each resource. This can also be done via a crowdsourcing approach. Research suggests that students as experts-in-training can accurately determine the quality of a learning resource and that the use of crowd-consensus algorithms in combination with optimal spot-checking by experts can increase the accuracy of assessment results (Whitehill, Aguerrebere & Hylak, 2019).

Not only can students create and evaluative resources effectively, but these activities also might enhance learning in and of itself. Classical and contemporary models of learning have emphasised the benefits of engaging students in activities across many higher-level objectives of the cognitive domain in Bloom's Taxonomy (Anderson & Krathwohl, 2001). In particular, students' development of creativity and evaluative judgment—"the capability to make decisions about the quality of work of self and others"—has been recognised as essential for student learning (Sadler, 2010). Honing these skills enables students to develop expertise in their field and to extend their understanding beyond their current work to future endeavors, including lifelong learning.

But can the vision of developing a cost-effective, discipline-agnostic AES via crowdsourcing be operationalised? An example of such a system follows.

RIPPLE: A CROWDSOURCED ADAPTIVE EDUCATIONAL SYSTEM

RiPPLE¹ is an adaptive learning system that recommends personalised learning activities to students, based on their knowledge state, from a pool of crowdsourced learning activities that are generated

¹ <https://ripplelearning.org/>

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

and evaluated by educators and the students themselves (Khosravi, Kitto, & Williams, 2019). RiPPLE integrates insights from crowdsourcing, learning sciences and adaptive learning, aiming to narrow the gap between these large bodies of research, and practical implementation into a platform that instructors can easily use in their courses. Figure 1 demonstrates an overview of one of the main pages of RiPPLE.

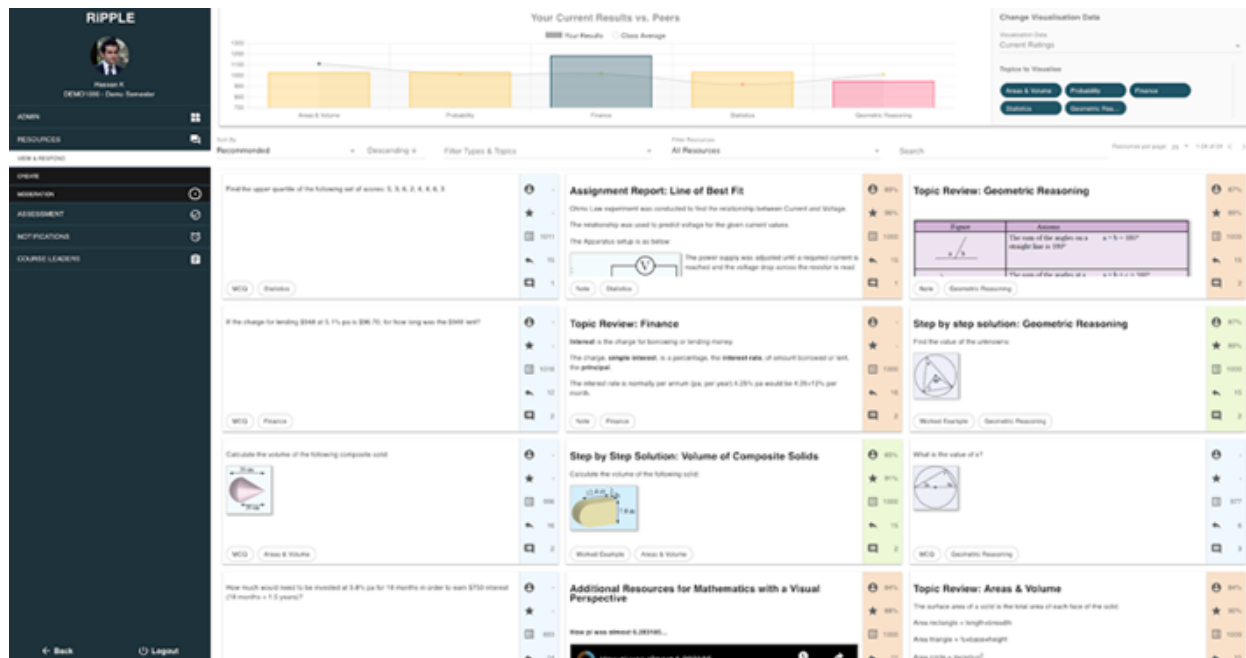


Figure 1: Overview of one of the main pages of RiPPLE

RiPPLE has the following three interconnected functions.

Student Modelling and Recommendation. The upper part of Figure 1 contains an interactive visualisation widget allowing students to view an abstract representation of their knowledge state based on a set of topics associated with a course offering. The colour of the bars, determined by the underlying algorithm modelling the student, categorises competence into three levels for a particular unit of knowledge red, yellow, and blue which signify, respectively, inadequate competence, adequate competence with room for improvement, and mastery. Currently, RiPPLE employs the Elo rating system for approximating the knowledge state of users (Abdi, Khosravi, Sadiq, Gasevic, 2019). The lower part of Figure 1 displays learning resources recommended to a student based on their learning needs.

Content creation. RiPPLE enables students to create a wide range of learning resources, including MCQs, worked examples, and general notes, incorporating text, tables, images, videos and scientific formulas. Given that students are developing as domain experts, it is likely that some of these learning resources may be ineffective, inappropriate or incorrect (Bates, 2014). Hence, there is a need for a moderation process to identify the quality of each resource. Here again, RiPPLE relies on the wisdom of the crowd and seeks help from students as moderators.

Content moderation: RiPPLE provides two “formal” moderation options that enable instructors to partner with students to review the quality of the student-created exercises before they are added

to a course's repository of learning resources. In both, (1) instructors determine the minimum number of moderations required per resource (e.g., 3 or 5) and (2) students review resources and provide a simple judgement, alongside a rationale for their decision. The two moderation options differ as to how the outcome of the process is determined. The two possibilities are (1) instructor's make the final call based on students' moderations or (2) the system automatically makes the final call based on students' moderation ratings and their reliability as computed by the system itself.

RiPPLE has the following three interconnected functions.

Student Modelling and Recommendation. The upper part of Figure 1 contains an interactive visualisation widget allowing students to view an abstract representation of their knowledge state based on a set of topics associated with a course offering. The colour of the bars, determined by the underlying algorithm modelling the student, categorises competence into three levels for a particular unit of knowledge red, yellow, and blue which signify, respectively, inadequate competence, adequate competence with room for improvement, and mastery. Currently, RiPPLE employs the Elo rating system for approximating the knowledge state of users (Abdi, Khosravi, Sadiq, Gasevic, 2019). The lower part of Figure 1 displays learning resources recommended to a student based on their learning needs.

Content creation. RiPPLE enables students to create a wide range of learning resources, including MCQs, worked examples, and general notes, incorporating text, tables, images, videos and scientific formulas. Given that students are developing as domain experts, it is likely that some of these learning resources may be ineffective, inappropriate or incorrect (Bates, 2014). Hence, there is a need for a moderation process to identify the quality of each resource. Here again, RiPPLE relies on the wisdom of the crowd and seeks help from students as moderators.

Content moderation: RiPPLE provides two “formal” moderation options that enable instructors to partner with students to review the quality of the student-created exercises before they are added to a course's repository of learning resources. In both, (1) instructors determine the minimum number of moderations required per resource (e.g., 3 or 5) and (2) students review resources and provide a simple judgement, alongside a rationale for their decision. The two moderation options differ as to how the outcome of the process is determined. The two possibilities are (1) instructor's make the final call based on students' moderations or (2) the system automatically makes the final call based on students' moderation ratings and their reliability as computed by the system itself.

To date, more than 5,000 registered users from 20 courses from 8 different disciplines have used RiPPLE to create over 8,000 learning resources and either attempt or review over 450,000 learning resources. In alignment with the literature, our findings suggest the following:

- Using RiPPLE as an AES that engages students in the creation and evaluation of resources led to measurable learning gains and, importantly, was perceived by students as beneficially supporting their learning (Khosravi, Kitto, & Williams, 2019).
- Providing open and transparent learner models as part of an AES can help students better understand their own learning needs and improve self-regulation (Abdi, Khosravi, Sadiq, & Gasevic, 2020).

- Using RiPPLE allows for the provision of personalised recommendations based on students' knowledge gaps and interests (Khosravi, Cooper & Kitto, 2017; Abdi, Khosravi, & Sadiq, 2019).
- Providing guides, exemplars, and rubrics supports students in developing their capacity for creating and evaluating resources—leading to an increase in the quality of the content repository. (Khosravi, Gyamfi, Hanna & Lodge, 2020)
- Utilising mechanisms such as gamification in education motivates students to be actively engaged, which can improve learning (Borges, Durelli, Reis & Isotani, 2014).
- Considering learning theories and pedagogical approaches is important for developing educational technologies; however, other factors such as usability, flexibility, and scalability are also critical (Khosravi, Sadiq & Gasevic, 2020).

PRACTICAL AND INTELLECTUAL CHALLENGES

While there exists adequate theoretical evidence regarding the potential of using crowdsourcing for development of AESs, there are many practical and intellectual challenges that still need to be addressed before this vision can be effectively operationalised at scale. A few of these challenges are listed below.

Quality Control. What mechanisms can crowdsourced AESs use to accurately judge the quality of a learning resource that is created by a student? Is the resource correct? Does it effectively help other students learn? Is it too similar to other resources that might have already been included in the resource repository?

Reliability Systems. How can crowdsourced AESs transparently, fairly, and accurately rate the reliability of each of the students?

Optimal Spot Checking. How can crowdsourced AESs optimally utilise the minimal availability of instructors in moderating resources to maximise the accuracy of the moderation process and reliability of student ratings?

Incentives. Despite students' personal beliefs and strong evidence from the learning science literature about the benefits of engaging in resource creation and moderation, based on our experience, students often require additional incentive mechanisms to engage with these activities. How can crowdsourced AESs incentivise students to engage with content creation and moderation?

Training and Support. Despite the recognition of the value of evaluative judgement and creativity in higher education, little attention has been paid to the development of tools and strategies to support their growth. How can crowdsourced AESs help students actively develop their creativity and evaluative judgment skills while creating learning resources?

Benchmark and Metrics. What benchmarking metrics can be used to measure the effectiveness of a crowdsourced AES?

Ethics: Ethical considerations should drive the design and implementation of crowdsourced AES. How can ensure that crowdsourced AESs comply with the ethical guidelines, protocols and principles which have been proposed by the learning analytics community?

REFERENCES

- Abdi, S., Khosravi, H., Sadiq, S., & Gasevic, D. (2020). A Multivariate ELO-based Learner Model for Adaptive Educational Systems. In the proceedings of the 12th International Conference on Educational Data Mining 12, 228 – 233
- Abdi, S., Khosravi, H., Sadiq, S., & Gasevic, D. (2020). Complementing Educational Recommender Systems with Open Learner Models. To appear in the proceedings of the 10th International Conference in Learning Analytics and Knowledge.
- Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. (2006, June). The cognitive tutor authoring tools (CTAT): preliminary evaluation of efficiency gains. In *International Conference on Intelligent Tutoring Systems* (pp. 61-70). Springer, Berlin, Heidelberg.
- Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2016). Instruction based on adaptive learning technologies. *Handbook of research on learning and instruction*. Routledge.
- Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. *Science*, 228(4698), 456-462.
- Anderson, L. W., Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives: Complete edition*. New York, NY: Longman.
- Bates, S. P., Galloway, R. K., Riise, J., & Homer, D. (2014). Assessing the quality of a student-generated question repository. *Physical Review Special Topics-Physics Education Research*, 10(2), 020105.
- Denny, P., Hamer, J., & Luxton-Reilly, A. (2009). Students sharing and evaluating MCQs in a large first year engineering course. In 20th Annual Conference for the Australasian Association for Engineering Education (AAEE '09), 6–9 December 2009, The University of Adelaide, Adelaide, Australia (pp. 575–580). Barton, Australia: Engineers Australia.
- De Sousa Borges, S., Durelli, V. H., Reis, H. M., & Isotani, S. (2014, March). A systematic mapping on gamification applied to education. In *Proceedings of the 29th annual ACM symposium on applied computing* (pp. 216-222). ACM.
- Essa, A. (2016). A possible future for next generation adaptive learning systems. *Smart Learning Environments*, 3(1), 16. <https://dx.doi.org/10.1186/s40561-016-0038-y>
- Galloway, K. W., & Burns, S. (2015). Doing it for themselves: students creating a high quality peer-learning environment. *Chemistry Education Research and Practice*, 16(1), 82-92.
- Khosravi, H., & Cooper, K., Kitto, K. (2017). RiPLE: Recommendation in Peer-Learning Environments Based on Knowledge Gaps and Interests. *Journal of Educational Data Mining*, 9(1), 42–67.
- Khosravi, H., Kitto, K., & Williams, J. J. (2019). RiPPL: A Crowdsourced Adaptive Platform for Recommendation of Learning Activities. *Journal of Learning Analytics*, 6(3), 83–97. <http://dx.doi.org/10.18608/jla.2019.63.12>
- Khosravi, H., Gyamfi, G., Hanna, B. & Lodge, J. (2010). Fostering and Supporting Empirical Research on EvaluativeJudgement via a Crowdsourced Adaptive Learning System. To appear in the Proceedings of the 10th International Conference on Learning Anytics and Knowledge.
- Khosravi, H., Sadiq, S., Gasevic, D., (2020) Development and Adoption of an Adaptive Learning System: Reflections and Lessons Learned. To appear in the proceedings of 51 Special Interest Group on Computer Science Education
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of educational psychology*, 106(4), 901.

- Nathan, M. J., Koedinger, K. R., & Alibali, M. W. (2001). Expert blind spot: When content knowledge eclipses pedagogical content knowledge. In Proceedings of the third international conference on cognitive science (Vol. 644648).
- Neil Heffernan, Korinn Ostrow, Kim Kelly, Douglas Selent, Eric Van Inwegen, Xiaolu Xiong, and Joseph Jay Williams. (2016) The Future of Adaptive Learning: Does the Crowd Hold the Key?, *International Journal of Artificial Intelligence in Education* 26 (2), 615-644.
- Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35(5), 535-550.
- Tackett, S., Raymond, M., Desai, R., Haist, S. A., Morales, A., Gaglani, S., & Clyman, S. G. (2018). Crowdsourcing for assessment items to support adaptive learning. *Medical teacher*, 40(8), 838-841.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197-221.
- Walsh, J. L., Harris, B. H., Denny, P., & Smith, P. (2018). Formative student-authored question bank: Perceptions, question quality and association with summative performance. *Postgraduate Medical Journal*, 94(1108), 97–103. <https://dx.doi.org/10.1136/postgradmedj-2017-135018>
- Whitehill, J., Aguerrebere, C., & Hylak, B. (2019). Do learners know what's good for them? Crowdsourcing subjective ratings of OERs to predict learning gains. *Educational Data Mining*, 2019, 12th.

The Modeling of Digital Learning Networks

Oleksandra Poquet

University of South Australia
sspoquet@gmail.com

Tobias Hecking

Universität Duisburg-Essen
tobias.hecking@uni-due.de

Bodong Chen

University of Minnesota
chenbd@umn.edu

ABSTRACT: Interest in social network analysis (SNA) as a way to gain insights into learning has existed in learning analytics (LA) since its inception. As a result, analytical approaches that harness the power of networks are playing important roles in the field. As models, networks allow us to visually communicate learning patterns. As a set of computational techniques, network analysis affords methods to generate network metrics for capturing learning. Despite the uptake of network analysis methods, we observe an underwhelming representation of their applications that (a) accommodate heterogeneous data of learning, (b) model network dynamics and network formation mechanisms within learning settings, or (c) derive new network metrics composite of heterogeneous information underlying network dynamics. The workshop we propose aims to gather LA scholars working in these areas to collectively build a foundation of advanced network modeling of learning data and shape strategies of future work in this important sub-field of LA.

Keywords: Network analysis; Network modeling; Social networks; Complex networks.

1 WORKSHOP ORGANIZERS

Sasha Poquet is a research fellow at the Centre for Change and Complexity in Learning (C3L), University of South Australia. She is interested in leveraging *peer effects* and *complex contagion* processes in the context of digital learning. Sasha researches how socio-technical networks form in digital settings, how to capture the impact of pedagogical interventions on network formation, and how to facilitate peer effects in digital learning environments.

Tobias Hecking is a postdoctoral researcher at the Department of Computer Science and Applied Cognitive Science at the University of Duisburg-Essen. His main research interest is on the development of Social Networks Analysis methods for understanding the production, acquisition, and dissemination of knowledge and innovation in digital settings. This includes information diffusion in online media, collaborative learning, as well as computer mediated collaboration in teams.

Bodong Chen is an associate professor and Co-Director of the Learning Informatics Lab at the University of Minnesota–Twin Cities. His research focuses on developing digital environments and analytics for collaborative learning and higher-order competencies. His recent work on learning analytics includes graph metrics for knowledge building, relational event models of online discussions, frequent sequence mining in MOOCs, and value-sensitive analytics design.

2 WORKSHOP/TUTORIAL BACKGROUND

The workshop builds on a long-standing interest of the LA community in social network analysis (SNA) as well as network science in general. Since the inception of this field, network analysis has been a pillar analytical approach (Siemens, 2013). SNA has shown to be powerful in revealing structural patterns of learner interactions in forums and social media, providing insights into social and collaborative learning processes. SNA is also used to provide feedback about social relations to learners and teachers so that they could reflect on the communication structure and act accordingly (Chen, Chang, Ouyang, & Zhou, 2018; Dawson, Bakharia, & Heathcote, 2010).

The applications of network analysis methods go beyond representations of interpersonal relationships, as networks can be used to analyze diverse relational data collected from learning settings. Bringing together different kinds of network actors have expanded the repertoire of analytical approaches in LA. For example, the Knowledge Building Discourse Explorer (KBDeX) supports socio-semantic network analysis that bridges social linkages with discourse units based on semantic overlaps and co-occurrences in learner utterances, allowing researchers to capture central ideas and key contributors at a particular time point (Oshima, Oshima, & Matsuzawa, 2012). Epistemic Network Analysis (ENA) is another method that combines content analysis with a network representation to uncover epistemic views of groups and individuals (Shaffer, Collier, & Ruis, 2016).

However, significant gaps still persist in the adaptation of network analysis in LA. Although the field of network science is rapidly developing, more advanced techniques for network analysis and modeling are not well represented in LA, and educational research in general. A review of applications of SNA in computer-supported collaborative learning (CSCL) showed that SNA applications are limited to descriptive reporting of SNA results and one-mode networks of learners (despite CSCL's emphasis on artifacts) (Dado & Bodemer, 2017). More advanced network analysis approaches that involve multiple types of nodes and edges (e.g., Contractor, 2009) and inferential network modeling (e.g., Butts, 2008) are rare. To reach more holistic understanding of learning, the community needs to move towards integrating different analytical angles in network analysis (Hecking, Chounta, & Hoppe, 2016).

As a result of these gaps, the community lacks novel network metrics that capture higher order learning constructs that may require a composite of multiple network metrics or metrics from a multidimensional network. Too often learning analytics studies use basic centrality measures in ill-defined network representations to serve as proxies of learning constructs. Further, researchers do not fully explore structural characteristics of a network, for example, the network topology on a more global level, nor generative processes that can explain baseline network formation mechanisms. Exciting work can be done to construct graph/network models in light of learning theories, map learning constructs onto components of the network, conduct inferential modeling of network factors

that play a role in learning, and derive new network metrics. Creating such metrics so that they are theoretically aligned, scientifically rigorous, empirically tested, and actionable, is a scholarly challenge, as well as an exciting opportunity for the field.

To tackle these challenges, as well as to strengthen the community's literacy in network analysis, we propose the workshop to gather scholars who are at the forefront of applying nascent network analysis in LA and colleagues from other methodological traditions eager to connect. This workshop will help us collectively craft an agenda of advancing network analysis as a pillar methodological tradition of learning analytics. This workshop is envisioned to be the first of a workshop series at LAK. The first gathering will work towards a special section of the Journal of Learning Analytics.

3 ORGANISATIONAL DETAILS OF PROPOSED EVENT

Type of event: Mini-tracks/Symposium OR Interactive workshop session

Proposed schedule and duration: Full-day

Type of participation: Open participation, meaning any interested delegate may register to attend. To present, participants need to submit proposals for consideration.

The workshop activities that participants should expect: The workshop will be organized around thematic contributions to provide discussion foci for participation in small groups (see possible themes listed above). This interactive format, which was very successful at prior workshops organized by the team, aims to facilitate an inclusive and effective discussion on the day, and onwards. Specifically, the workshop will include the following components: approximately 5 paper presentations; groups organized around presentations; a "birds of a feather" activity organized around themes that emerge from discussions; large-group discussion that sets the agenda for future work in this area.

Expected participant numbers and planned dissemination activities to recruit attendants: 20-30 participants. The workshop organizers are embedded in the learning analytics and related communities. They will make use of listservs (SoLAR, Learning Analytics Google group, EDM-announce, ISLS/CSCL, AERA SIG-LS, EARLI) and leverage their own personal networks to advertise the workshop. We will especially reach out the Network Science (NetSci) and International Network for Social Network Analysis (INSNA) communities. Researchers, practitioners, and funders indicate an increasing interest in network modeling and learning analytics, and approaches to put network analytics into practice are currently at the forefront of many learning analytics efforts, thus we anticipate the workshop having popular appeal.

Required equipment for the workshop:

- Pinboard(s) __4__ piece(s)
- Post-its __8__ piece(s): 4 different colors (2 pieces for each color)
- Seating: table & chair positions can be changed; table groups for __5__ people each

4 WORKSHOP OBJECTIVES OR INTENDED OUTCOMES

The workshop objectives are: to explore the application of advanced network analysis and modeling to learning data; to explore means of connecting multiple analytical dimensions through network analysis; to brainstorm novel network metrics. The underlying goal is to build a sub-community of researchers to lead future development in this area. Accepted workshop papers will be published on the workshop website and in a joint LAK Companion Proceedings. Workshop outcomes will be disseminated on Twitter using #LAK20Network. The workshop website is set up using the Github Pages: <https://colig.github.io/lak20network/>. The website contains the call for workshop presentations, the workshop program, any shared sample datasets, codes, slides, and so on.

REFERENCES

- Butts, C. T. (2008). A relational event framework for social action. *Sociological Methodology*, 38(1), 155–200. <https://doi.org/10.1111/j.1467-9531.2008.00203.x>
- Chen, B., Chang, Y.-H., Ouyang, F., & Zhou, W. (2018). Fostering student engagement in online discussion through social learning analytics. *The Internet and Higher Education*, 37, 21–30. <https://doi.org/10.1016/j.iheduc.2017.12.002>
- Contractor, N. (2009). The Emergence of Multidimensional Networks. *Journal of Computer-Mediated Communication: JCMC*, 14(3), 743–747. <https://doi.org/10.1111/j.1083-6101.2009.01465.x>
- Dado, M., & Bodemer, D. (2017). A review of methodological applications of social network analysis in computer-supported collaborative learning. *Educational Research Review*, 22, 159–180.
- Dawson, S., Bakharia, A., & Heathcote, E. (2010). SNAPP: Realising the affordances of real-time SNA within networked learning environments. In L. Dirckinck-Holmfeld, V. Hodgson, C. Jones, M. de Laat, D. McConnell, & T. Ryberg (Eds.), *Proceedings of the 7th International Conference on Networked Learning* (pp. 125–133). Lancaster, UK: Lancaster University.
- Hecking, T., Chounta, I.-A., & Hoppe, H. U. (2016). Investigating social and semantic user roles in MOOC discussion forums. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 198–207. <https://doi.org/10.1145/2883851.2883924>
- Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: A social network analysis application for knowledge building discourse. *Educational Technology Research and Development: ETR & D*, 60(5), 903–921. <https://doi.org/10.1007/s11423-012-9265-2>
- Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A Tutorial on Epistemic Network Analysis: Analyzing the Structure of Connections in Cognitive, Social, and Interaction Data. *Journal of Learning Analytics*, 3(3), 9–45. <https://doi.org/10.18608/jla.2016.33.3>
- Siemens, G. (2013). Learning analytics: The emergence of a discipline. *The American Behavioral Scientist*, 57(10), 1380–1400.

Exploration of Peer Effects through Digital Forum Interactions

Oleksandra Poquet

Centre for Change and Complexity in Learning
University of South Australia
sspoquet@gmail.com

ABSTRACT: This submission presents a work-in-progress on the identification of peer effects using data collected in digital learning environments. One of the central claims in the economics of higher education is that peer exposure has potential to affect academic choices and educational attainment. Studies of peer effects using learning analytics data, however, have been scarce, and mainly focused on highlighting social selection, i.e. phenomenon when similar learners choose to interact with similar learners. This poster presents exploratory work that examines a different peer mechanism: social influence through online forum interactions where *a change* of learner attributes is associated with the attributes of peers they interacted with online. To examine if the change of learner performance is associated with the performance of the peers they interact with in online forum, we examined temporal development of forum interactions at the university-level collected during 2013-2016. The submission reports on preliminary results. The associations found in our analyses should be interpreted with caution as the mechanisms of selection and influence in blended and online environments are unlikely to be fully equivalent to those observed in traditional social networks.

Keywords: peer effects, digital networks, online forum, stochastic actor-oriented models

1 PROBLEM FORMULATION

One of the central claims in the economics of higher education is that peer exposure has potential to affect academic choices and educational attainment (Winston and Zimmerman, 2004). Peer effects are considered to exist when a person's behavior is affected due to the presence of peers with particular characteristics (ibid., p. 396). For instance, adding more female learners into peer groups can lead to the increase in scores in reading and math; whereas high school students from underprivileged groups that are exposed to low achieving cohorts, score lower themselves (Hoxby, 2000). Studies reviewing evidence of peer effects on scores reported them as modest-sized and statistically significant (Sacerdote, 2011). This suggests that interpersonal relationships developed during learner trajectories contribute to the learning outcomes.

Traditional peer effects research considers social friendship or direct face-to-face classroom and high-school interactions as a foundation of peer-to-peer relationships. These relationships are usually inferred from learner-reported data of peer networks. Only a handful of studies leveraged learning analytics approaches to explore if technology-mediated interactions may capture peer effects on one's attainment and learning. For example, learning analytics research using co-enrolment records (e.g. Gasevic, Zouq, and Janzen, 2013; Gardner, Brooks, Li, 2018) suggests that in both fully online and residential programs student performance can be, to some extent, predicted, using information about their course peers.

Research using peer interactions collected from online forums similarly suggests the relevance of peer attributes and patterns of communication to student learning. Learning analytics research on online forum interactions has shown that high performing students often interact with high performing students. An example is a study of online interactions by Huang & Chen (2018) who found that higher-prestige students formed a dense ‘rich club’ less likely to interact with a lower-prestige group. However, the mechanisms behind this pattern remain unclear. A pattern of a high achiever interacting with a high achiever can result from two distinct mechanisms: 1) social selection or homophily, where high achievers choose to interact with high achievers; 2) social influence or peer effects, where high performers may have influenced lower-achieving students who over time improved their performance. Temporal analyses show that often online forums garner social selection processes (Huang and Chen, 2018; Vaquero & Cebrian, 2013). Yet, to what extent they also contain the presence of peer effects, i.e. social influence processes, has not been explicitly addressed.

2 RESEARCH QUESTION

To explore if both social selection and social influence processes may be observed through digital forum interactions, this study inquired:

To what extent longitudinal networks of online forums interactions result from peer selection based on performance, and to what extent from peer influence?

This research question did not assume that learners impacted other learners’ grades, but rather that the association of changes in performance with previously occurred peer interactions could serve as evidence of the presence of peer effects. We assume that the specific mechanisms for social selection and influence are unobserved in our model, given that we only leveraged student log data to construct networks and their average grades at the semester level.

3 METHODS

Using forum interactions in an in-house online platform of a large XXX university, we constructed four networks representing online interactions from September 2013-February 2016, of the cohort that started in 2012. Student grades were averaged across each semester representative of the time slice. Average grades in the previous semester were used to predict interactions between peers in the subsequent semester; forum interactions in the previous semester were used to predict grades in the subsequent semester. To this end, stochastic actor-oriented models (SAOMs) were used (Snijders, 2011). SAOMs require a threshold of stability across data observations in time slices, which appears to be a challenge in the temporal forum interaction data. We have obtained the sufficient data overlap by modelling networks of students who had interactions across at least two semesters. Final model included interactions between 230 students, whereas their intermittent interactions with other students were added to the model as controls for the effect of less frequent digital encounters.

4 RESULTS

Convergence of the final model was 0.23, goodness of fit was acceptable. The model supports the growth of ties ($\beta=5$, $SE=0.1$, $p<0.001$) and reciprocity ($\beta=3$, $SE=0.1$, $p<0.001$) in the digital

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

interactions, as well as controls for tie formation propensity by both performers whose grade was in the lower 25% bracket ($\beta=0.6$, $SE=0.1$, $p<0.001$) and performers whose performance was in the highest 25% bracket ($\beta=0.2$, $SE=0.1$, $p>0.05$).

According to the results of this work-in-progress study, social selection processes were not statistically significant across the entire student population ($\beta=0.07$, $SE=0.69$). This differed for the performers whose grade was in the top 25% of the university ($\beta=0.57$, $SE=0.1$, $p<0.001$). Specifically, these high achievers tended over time to form clusters within online interaction.

The model shows that overall student performance as observed through the networks have improved ($\beta=1.23$, $SE=0.57$, $p<0.05$), in particular that was true for learners whose performance was in the higher brackets to start out with ($\beta=1.57$, $SE=0.6$, $p<0.05$). In relation to peer effects, we observed a strong significant effect of performance assimilation ($\beta=33.6$, $SE=3.7$, $p<0.001$). The model supported the tendency of learners' grades to assimilate with the grades of those they interacted with during their university study.

The findings describe online interactions that were sustained between two learners for at least two semesters. Yet, we are cautious to interpret the mechanisms for the observed effects. Our replication of the model in another context suggests that the patterns may be contingent on the country and university cultures. The poster will present the results of the final model, specify the directionality of peer effects, and discuss study's limitations.

REFERENCES

- Gardner, J., Brooks, C., & Li, W. (2018). Learn from Your (Markov) Neighbour: Co-Enrollment, Assortativity, and Grade Prediction in Undergraduate Courses. *Journal of Learning Analytics*, 5(3), 42-59.
- Gašević, D., Zouaq, A., & Janzen, R. (2013). "Choose your classmates, your GPA is at stake!" The association of cross-class social ties and academic performance. *American Behavioral Scientist*, 57(10), 1460-1479.
- Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation (No. w7867). National Bureau of Economic Research.
- Huang, T., & Chen, B. (2018). Uncovering the rich club phenomenon in an online class. In R. Luckin, & J. Kay (Eds.), *13th International Conference of the Learning Sciences, ICLS 2018: Rethinking Learning in the Digital Age: Making the Learning Sciences Count* (2018-June ed., Vol. 3, pp. 1375-1376). International Society of the Learning Sciences (ISLS).
- Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far?. In *Handbook of the Economics of Education* (Vol. 3, pp. 249-277). Elsevier.
- Snijders, T. A. (1996). Stochastic actor-oriented models for network change. *Journal of mathematical sociology*, 21(1-2), 149-172.
- Vaquero, L. M., & Cebrian, M. (2013). The rich club phenomenon in the classroom. *Scientific reports*, 3, 1174.
- Winston, G., & Zimmerman, D. (2004). Peer effects in higher education. In *College choices: The economics of where to go, when to go, and how to pay for it* (pp. 395-424). University of Chicago Press.

The Current State of Knowledge-Building Analytics and Possible Future Directions

Jun Oshima

RECLS, Shizuoka Univ.
joshima@inf.shizuoka.ac.jp

Ritsuko Oshima

RECLS, Shizuoka Univ.
roshima@inf.shizuoka.ac.jp

ABSTRACT: Knowledge-building is the most prominent theory in the metaphor of learning as knowledge-creation, and its pedagogical approach facilitates the development of CSCL systems such as Knowledge Forum and activates many design-based studies in the world. In knowledge-building, learners engage in improving their ideas by utilizing conceptual artifacts through collaborative discourse. Although qualitative analysis provides fine-grained pictures of knowledge building practices, the quantitative approach needs to be developed for handling extensive data and conducting more powerful analyses in the mixed-methods. In this study, we discuss the current state of the quantitative analysis of kb discourse from the socio-semantic network analysis and possible future directions with the development of algorithms and technologies.

Keywords: Knowledge-building analytics, socio-semantic network analysis, the mixed-methods approach.

1 THEORETICAL BACKGROUND

In knowledge-building, knowledge is considered as an object to improve continuously (Scardamalia & Bereiter, 2014). It is assumed that the knowledge is collective, and the community of people shares and discusses their ideas through discourse then improves knowledge used in their ideas. Every learner has the collective cognitive responsibility to engage in the knowledge-building discourse. Through knowledge-building discourse, learners engage in improving their ideas comprised of their knowledge by using available conceptual artifacts. In the case that educational researchers use statistics to test their hypotheses, the hypotheses themselves are ideas to improve. The statistics they use for examining their hypotheses is a conceptual artifact. Thus, we need to consider analytics to capture what ideas learners improve (content-oriented) and how they improve their ideas by using their available conceptual artifacts (epistemic practice-oriented) to evaluate knowledge-building discourse.

2 THE CURRENT STATE OF KNOWLEDGE-BUILDING ANALYTICS

In the knowledge-building research community, several researchers have used a socio-semantic network analysis such as Knowledge Building Discourse Explorer (KBDeX). In the socio-semantic network analysis, they attempt to figure out how students engage in their collective knowledge advancement (e.g., Oshima et al., 2012), rotate their leaderships for improving their ideas (Ma et al.,

2016), and exert their shared epistemic agency through collaborative discourse (Oshima et al., 2018). Moreover, some research groups have developed new tools to evaluate idea promisingness (Chen et al., 2015) based on written discourse in notes on Knowledge Forum (Lee & Tan, 2017), and to provide learners with formative feedback for them to consider how they can extend their ideas (Feng et al., 2019).

While the socio-semantic network analysis of vocabulary has provided researchers with new insights on knowledge-building discourse, researchers have not fully discussed the epistemic practices in the knowledge-building discourse, i.e., how learners improve their ideas through their collaborative discourse. For examining the epistemic practice, Shaffer et al. (2017) proposed Epistemic Network Analysis (ENA). Their approach was also based on the connection between elements in the discourse, but more practice-oriented, i.e., codes representing cultural practices. ENA relies on the epistemic frame theory.

In their epistemic frame theory, Shaffer and colleagues (Rohde and Shaffer 2004) presumed that we use unique grammar in an established community of practice. First, we as humans have our epistemic frames formed as a collection of skills, knowledge, identity, value, and epistemology in the cultural grammar. Second, we gradually internalize the epistemic frames through participation in community practices. Third, we use the epistemic frame of a community when the specific perspective of a community determines how we act.

For conducting ENA, researchers have to identify what components could be detected to represent the epistemic frame. The components are Codes of the cultural practice in a community. They code the presence of cultural codes in each discourse analysis unit, then create network structures of the Codes by ENA based on the adjacency matrix of the co-occurrence of the Codes (Shaffer, 2017). Codes are plotted on a two-dimensional space created through a multivariate statistical procedure similar to factor analysis with varimax rotation (Shaffer, Collier, & Ruis, 2016). The statistically meaningful space for Codes makes it possible for researchers to directly compare the epistemic frames between different groups (e.g., high and low learning-outcome groups). The authors examined the shared epistemic agency by high-school students in their small group works by using ENA. They defined seven epistemic actions identified by Damşa et al. (2010) as Codes for identifying differences in their epistemic frames between high and low learning-outcome groups as well as another analysis of idea improvement process by using KBDeX. The double-layered analysis of discourse by ENA and KBDeX provided researchers with a more accurate understanding of differences in knowledge-building discourses between successful and unsuccessful groups. It was found that successful students controlled their epistemic actions to produce more ideas in their early stages of learning compared with unsuccessful ones.

3 FUTURE DIRECTIONS OF KNOWLEDGE-BUILDING ANALYTICS

In the presentation at this workshop, the authors summarize the current state of knowledge-building analytics and propose some future directions considering key concepts behind the socio-semantic network analysis such as KBDeX and ENA. First, we like to discuss the unit of analysis in more detail. Researchers rely on the co-occurrence of elements for creating a network graph. The co-occurrence is dependent on what unit of analysis we are going to define. In the analysis of discourse, dialogue is a key concept. What is the smallest unit of the dialogue? We will discuss it in

the session. Second, the current state of knowledge-building analytics relies on discourse data. However, nonverbal acts may influence group dynamics. A new direction of knowledge-building analytics would be directed at the multimodality of datasets. What kinds of action logs may be needed for further analysis? We also will discuss it in the session.

REFERENCES

- Chen, B., Scardamalia, M., and Bereiter, C. (2015). Advancing knowledge-building discourse through judgments of promising ideas. *International Journal of Computer-Supported Collaborative Learning*, 10, 345–366.
- Damşa, C. I., Kirschner, P. A., Andriessen, J. E. B., Erkens, G., & Sins, P. H. M. (2010). Shared epistemic agency – An empirical study of an emergent construct. *Journal of the Learning Sciences*, 19(2), 143–186.
- Feng, X., van Aalst, J., Chan, C. K. K., Yang, Y. (2019). Fostering collective knowledge advancement through idea-friend maps in a large knowledge building community. In Lund, K., Niccolai, G., Lavoué, E., Hmelo-Silver, C., Gweon, G., and Baker, M. (Eds.), *13th International Conference on Computer Supported Collaborative Learning (CSCL) 2019, Volume 1* (pp. 288–295). Lyon, France: International Society of the Learning Sciences.
- Lee, A. V. Y., & Tan, S. C. (2017). Promising Ideas for Collective Advancement of Communal Knowledge Using Temporal Analytics and Cluster Analysis. *Journal of Learning Analytics*, 4(3), 76–101.
- Ma, L., Matsuzawa, Y., Chen, B., & Scardamalia, M. (2016). Community Knowledge, Collective Responsibility: The Emergence of Rotating Leadership in Three Knowledge Building Communities. In C. K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), *The International Conference of the Learning Sciences (ICLS) 2016, Volume 1*. Singapore: International Society of the Learning Sciences.
- Oshima, J., Oshima, R., & Fujita, W. (2018). A Mixed-Methods Approach to Analyze Shared Epistemic Agency in Jigsaw Instruction at Multiple Scales of Temporality. *Journal of Learning Analytics*, 5(1), 10–24.
- Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: A social network analysis application for knowledge building discourse. *Educational Technology Research & Development*, 60, 903–921.
- Rohde, M., and D. W. Shaffer. (2004). Us, ourselves, and we: Thoughts about social (self-) categorization. *Association for Computing Machinery (ACM) SigGROUP Bulletin*, 24 (3), 19–24.
- Scardamalia, M. & Bereiter, C. (2014). Knowledge building and knowledge creation: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences: Second edition* (pp. 397–417). New York, NY: Cambridge University Press.
- Shaffer, D. W. (2017). *Quantitative ethnography*. Madison, WI: Cathcart Press.
- Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3(3), 9–45.

Modelling students' social network structure from spatial-temporal network data

Quan Nguyen

School of Information, University of Michigan
quanngu@umich.edu

Warren Li

School of Information, University of Michigan
liwarren@umich.edu

Christopher Brooks

School of Information, University of Michigan
brooksch@umich.edu

ABSTRACT: Network analysis in educational research has primarily relied on self-report and/or data generated from online learning environments (e.g. discussion forums). However, a large part of students' social connections occurs through day-to-day interactions on campus. This paper describes an on-going work exploring the application of WiFi network data to model social network structure amongst students on campus at scale. Links between individuals were inferred based on their spatial co-occurrences along a similar temporal dimension (i.e. two individuals connected to the same WiFi access point at the same time). We discussed a potential approach to test the statistical significance of these connections against a null model, in which two individuals might randomly be at the same place at the same time.

Keywords: Network analysis, spatial-temporal data, WiFi log data

1 WIFI NETWORK DATA IN EDUCATION

Wireless local area networks (WLANs) have increasingly become ubiquitous in modern education as they provide seamless internet access to students, teachers, and staffs through a large number of WiFi access points on campus. Log-files generated from WLANs are rich in both temporal and spatial features. They recorded the timestamp of each user's devices being connected to a particular WiFi access point (Table 1).

Table 1: Anonymized sample WiFi data.

User ID	Timestamp	Access point	MAC address
A1234	2018-09-24 08:00:00	TWC-1023NW	Android- A1234
A1234	2018-09-24 08:02:03	TWC-2013NW	Android- A1234
B2314	2018-09-24 08:00:03	BAHR-1210-N	Apple- B2314
C2153	2018-09-24 08:00:05	CQTB-3734	Ubuntu- C2153

While there has been extensive research using WiFi data focusing on signal processing, only a limited number of studies has explored the application of WiFi data for educational purposes. For example, the iSpots project at MIT collected WiFi data and visualized the dynamics changes in wireless traffic on the wireless network and showed how people move around campus in real-time (Sevtsuk, 2009). WiFi data has also been used in predictive modelling. Sarkar, Carpenter, Bader-El-Den, and Knight (2016) estimated the correlations between students' time spent on campus based on WiFi log data and academic performance. In another study, Hang, Pytlarz, and Neville (2018) combined WiFi log data with building location profiles at Purdue University to extrapolate the temporal dynamics of user's location preferences throughout the day, and to predict *Point of Interest* (POI) (e.g. where an user will be on Monday at 9:00 am) using graph embeddings. For instance, Zhou et al. (2016) utilized WLAN data at Tsinghua University to estimate students' punctuality (attendances, late arrivals, and early departures) for lectures as well as to assess the lecture's engagement using mobile phone's interactive states at minute-scale granularity. However, due to the sensitive nature of WiFi data, researchers should be cautious and transparent about their purposes.

Another promising application of WiFi data in educational research, which has yet to be explored, is to understand the social network structure of students. WiFi network data can help researchers capture the dynamic changes in social interactions on campus, which in turns, can be combined with discussion forum data and self-report social network surveys. Nonetheless, the spatial-temporal nature of WiFi data presents unique conceptual and methodological challenges for network analysis, which will be discussed below.

2 NETWORK INFERENCE FROM SPATIAL-TEMPORAL DATA

Data in this study were collected from 3,915 students enrolled in five large undergraduate STEM courses at a public university in the U.S. in the Fall semester of 2018. All students' identifiers were anonymized. To get a sense of the data, Figure 1 visualizes the temporal changes in WiFi access points of two users on a particular date from 08:00 to 20:00. These two users spent a large amount of time in the morning at a fixed WiFi access point, possibly attending a lecture. In the afternoon, these two users shared the same location for 2 hours. After that, each user went on about their day to different areas on campus.

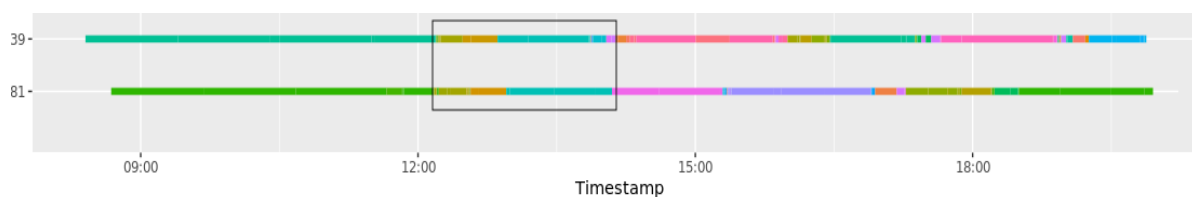
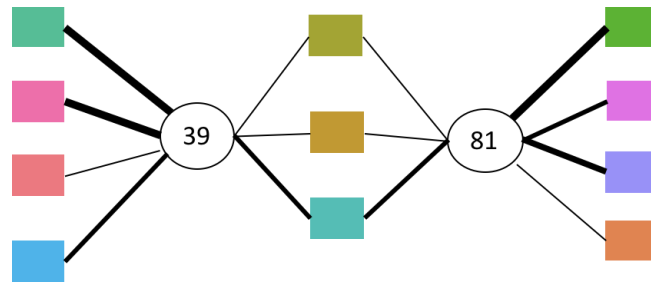


Figure 1: Temporal changes in WiFi access points of two users throughout a day. The boxed area indicates a two-hour period where these users shared the same access point

WiFi log data can be treated as a bipartite network (i.e. user-access point) along a temporal dimension (Figure 2). This can be projected into an undirected weighted one-mode network (i.e. user-user) under the following assumptions:

- Two nodes are linked if they connected to the same WiFi access point within the same time window (i.e. to be at the same place at the same time) (Figure 1).
- Tie's weight is determined as the shared duration for the same WiFi access point
- Tie's weight is discounted for the number of nodes sharing the same access point (i.e. the more people in the room, the weaker the tie between two particular individuals)
- Tie's connections could occur by random chance
- Tie's connections could occur due to shared events (e.g. attend the same lecture)



**Figure 2: Bipartite network of two users (circles) and WiFi access points (squares).
Tie's weight/thickness represents connected duration.**

Based on this, we can calculate the total amount of shared duration between two particular users on a given day and aggregate them across a week/semester/year to create a weighted adjacency matrix amongst users. To test the statistical significance of ties, we propose drawing a random observation for each user (i.e. random amount of time spent at a random WiFi access point at a random time of day). This permutation can be repeated to generate a null distribution of shared duration between two particular users, which allow us to test the statistical significance of a given tie (Psorakis, Roberts, Rezek, & Sheldon, 2012). Due to the limited space and the early stage of this work, more concrete results will be presented and discussed at the conference. Nonetheless, we believe the unique opportunities, as well as challenges of using spatial-temporal data for network analysis in education, would bring out a lot of interesting discussions.

REFERENCES

- Hang, M., Pytlarz, I., & Neville, J. (2018). *Exploring Student Check-In Behavior for Improved Point-of-Interest Prediction*. Paper presented at the Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, United Kingdom.
- Psorakis, I., Roberts, S. J., Rezek, I., & Sheldon, B. C. (2012). Inferring social network structure in ecological systems from spatio-temporal data streams. *Journal of The Royal Society Interface*, 9(76), 3055-3066.
- Sarkar, S., Carpenter, B., Bader-El-Den, M., & Knight, A. (2016, 6-8 July 2016). *Where students go and how they do: Wi-Fi location data versus academic performance*. Paper presented at the 2016 9th International Conference on Human System Interactions (HSI).
- Sevtsuk, A. (2009). Mapping the MIT campus in real time using WiFi *Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City* (pp. 326-338): IGI Global.
- Zhou, M., Ma, M., Zhang, Y., Sui, A., K., Pei, D., & Moscibroda, T. (2016). *EDUM: classroom education measurements via large-scale WiFi networks*. Paper presented at the Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany.

Tree Structure of Collective Attention Network: Revisiting the Problem of Dropout

Jingjing Zhang

School of Educational Technology, Beijing Normal University

Jingjing.zhang@bnu.edu.cn

Ming Gao

Research Centre of Distance Education, Beijing Normal University

mgao519@mail.bnu.edu.cn

ABSTRACT: Treating the dropout phenomenon as a sign of an individual's choice highlights the importance of understanding how dropouts learn in MOOCs. Conventional learning analytics methods failed to make sense of limited behavior data left by dropouts. This study uses the minimum spanning tree of collective attention network to investigate how dropouts behave in a selected MOOC. It is interesting to note that assessments embedded in the MOOCs seem to play a rather important role in guiding dropouts to learn. Redefining assessment in open and flexible learning environments to construct a minimum cost network of collective attention is vital to make this online space cost-effective for better learning.

Keywords: MOOCs, dropout, collective attention, learning analytics, pattern mining

1 INTRODUCTION

Alongside the increasing development of MOOCs accommodating open and flexible learning experiences at scale, the unusual high dropout rates beyond 90% of participations are alarming (Jordan, 2014). The dropout phenomenon was first treated as a sign of deficient quality, but later as an explicit expression for an individual's choice. This later counter-argument highlights the importance of focusing on how dropouts learn in MOOCs, and in what ways the learning design of MOOCs facilitates their learning. Nevertheless, few studies have taken the learning analytics approach to understanding how dropouts learn, as no or little behavior data of dropouts can be meaningfully addressed using the data mining approach. Furthermore, the behavior data of dropouts are often removed as "outliers" in traditional statistical analysis. To address such a problem, we build upon the earlier research by using the network model of collective attention to investigate how dropouts learn at the collective level in a MOOC. A key innovation is the focus on how to make sense of learning patterns by using a new method to model short, limited, and heterogeneous behavior trajectories left by dropouts.

2 METHODS

2.1 Context

'Introduction to Psychology (2018 autumn)' offered on XuetangX was selected as the case to study. This course offered 70 learning resources, including videos, quizzes, and an exam, within 13 units. 9508 learners participated in the course, and their behavior data were automatically stored in the

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

database. Due to the incomplete records of behavior data (no. 5237) via mobile devices, deficient information of registration and exam (no. 2110), only 1892 dropouts out of 2161 were selected for this study. The dropout rate of 88% of this course is a typical rate frequently reported in the literature. About 200 dropouts registered before the course started, and about 400 dropouts registered after the course ended. The accumulated number of participations increased over time, and half of the learners have registered before the mid-term. This pattern of registration reflects on learners' choice to learn at their pace. The pattern of learners' visits is also a typical long-tail distribution. While about 700 participants accessed courseware in unit 1, the learners drop out over time, resulting in only 16 learners accessed the last unit – exam.

2.2 An Open-Flow Network of Collective Attention and its Minimum Spanning Tree

The classical social network is a closed model that fails to account for the high rates of attrition and steeply unequal participation patterns of learners. This study built upon our earlier research, and adopted the open flow network of collective attention (Zhang, Lou, Zhang, & Zhang, 2019) to model learners behaviors, and see this article for a comprehensive review of collective attention. This network model, using node to represent the learning resources and the link to represent the learners' sequential visits across the learning resources. At the collective level, the large body of learner's sequential visits resembles the flux of attention flowing in and out of the learning resources. The flux of such attention flow forms a network, in which two artificial nodes - 'source' and 'sink'- were added to represent the offline space. Thus, this network becomes an open and balanced model, which allows collective attention to flow in and out across online and offline spaces. As for individual learning resources, the inflow of attention equals the outflow.

In such a collective attention network, flow distance (Guo et al., 2015) measures the average first-arrival distance between nodes, by using the N-order Markov transition to calculate the probability that attention would flow in or out of a learning resource. A skeleton of the network, including all learning resources, was generated by using Minimum Spanning Tree (MST) (Kruskal, 1956). In such a tree structure, for any node, another node with the shortest flow distance to it was added until all the nodes were added to the tree containing a sum of flow distances, which is minimum. Zhang and her colleagues (2019) found that the amount of attention flow and flow distance were negatively correlated. Thus, the weight of link is calculated using the opposite of flow distance between two nodes, which implies that the likelihood that amount the attention flow (including direct and indirect) in and flow out between two nodes in such a network. The learning resources in the same color belong to the same unit, and the size of the node is proportional to the amount of attention flow in/out to this learning resource. Python and Gephi were used for data analysis and visualization.

3 RESULTS AND CONCLUSION

The minimum spanning tree represents a new structure formed by using real behavior data of dropouts (see Figure 1). Such a tree, representing the topological properties of the collective attention network, yields a lower bound on the cost of collective attention. As argued by Zhang et al. (2019), MOOC learning is pricey at the cost of the learners' attention, and thus the topological structure of such a tree contributes to the design of cost-effective learning resources to prevent learners from becoming overloaded.

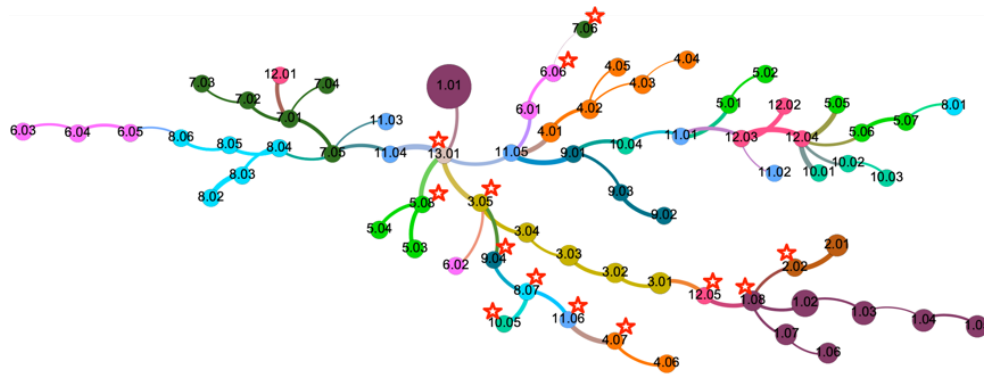


Figure 1: The minimum spanning tree of learners' collective attention flow

The visiting pattern of dropouts presents a long-tail distribution, which also reflects on the MST. As shown in Figure 1, the size of nodes belonging to Unit 1 is the largest, while the size of node 13.01, the exam, is the smallest. This illustrates that the amount of attention flow decreases from earlier units to later units. However, it is interesting to discover that the largest node 1.01, which the most of collective attention flows in and out, is not the center of this MST. It is instead an isolated leave connecting to the central node of the exam, and the rest learning resources belonging to Unit 1 are clustered together lie leaves on the other side. Instead, the exam Unit (13.01) is the distinct center of the MST, which implies that the cost of collective attention is minimum by giving the assessment a central role to connect with other learning resources.

We can also see in this tree a separation of the learning resources into three large branches of resources and several twigs of varied lengths. It is interesting to note that quizzes are likely to be at the crossing of the main branches, such as quiz 1.08, 3.05, 5.08, and 8.07 (marked in red star in Figure 1). Notably, several quizzes across different units (e.g., 3.05, 4.07, 8.07, 9.04, 10.05, 11.06) form a cluster, and quizzes are also likely to act as the bridges between video resources. For example, Quizzes in unit 1, 2 and 12 serve as the bridges to link unit 1, 2, and 3. One possible explanation for this result is that quizzes are not used by dropouts to evaluate their studies, instead dropouts use quizzes as a learning strategy to guide their study.

This preliminary exploration of how dropouts learn in a selected MOOC only sheds light on behavior patterns using the model of collective attention and its MTS. Likely learning intentions and learners' profiles, as argued in the literature, play a significant role in constructing similarities in patterns of learning behaviors. In our future work, how to incorporate nonstructural properties, such as intention, capacity, time, etc. in the model of collective attention is to be seriously considered.

REFERENCES

- Jordan, K. (2014). Initial trends in enrolment and completion of massive open online courses Massive Open Online Courses. *International Review of Research in Open and Distance Learning*, 15(1).
- Kruskal, J. B. (1956). On the Shortest Spanning Subtree of a Graph and the Traveling Salesman Problem. *Proceedings of the American Mathematical Society*, 7(1), 48.
- Guo, L., Lou, X., Shi, P., Wang, J., Huang, X., & Zhang, J. (2015). Flow distances on open flow networks. *Physica A: Statistical Mechanics and Its Applications*, 437, 235–248.
- Zhang, J., Lou, X., Zhang, H., & Zhang, J. (2019). Modeling collective attention in online and flexible learning environments. *Distance Education*, 00(00), 1–24.

The Flow of Books in the Era of Social Media: A case study of a reading group

Lei Zhang

Department of Education Information Technology, East China Normal University
azhangyilei@163.com

Xiangdong Chen

Department of Education Information Technology, East China Normal University
xdchen@deit.ecnu.edu.cn

ABSTRACT: Although many researchers have investigated individual factors (e.g. motivation, price of book, etc.) impacting on readers' book choice, few research has been done to investigate the social factors impacting on people's book choice in the era of social media, in which more and more interact with each other. The purpose of the study was to explore learners' book choice from the perspective of social network, and provide results for those who are interested in the flow of books under social media spaces. A relatively closed reading group, including 42 readers, was selected to be surveyed and interviewed. Characters of this reading group, readers' roles, and factors influencing book choice were analyzed and represented. Issues and recommendations about the flow of books were discussed.

Keywords: social reading, book choice, social network

1 INTRODUCTION

Group reading, or social reading, means the process in which a group of readers share reading feelings and receive feedback (Dean, 2016; Vlieghe, Vandermeersche, & Soetaert, 2016). With the emergence of social media devices (e.g. Facebook, twitter), a great number of readers are influenced by other readers through social media when selecting a book. However, few studies explored readers' book choice from the perspective of social network. Although there were some research investigating factors that influenced readers' book selection (Bang-Jensen, 2010), there existed some limitations. One limitation is the lack of investigation of social factors influencing readers' book choice. Although those studies mentioned the phenomenon that readers' book selection can be influenced by social interactions with others, deep inspection for the reasons was not conducted. The purpose of this article is to investigate social features of a reading group, investigate social factors influencing readers' book choice from the perspective of social network, and propose suggestion for people who are interested in the flow of books in the era of social media.

2 METHODOLOGY

A total of 42 first-undergraduate students (13 males, 29 females) in Educational Technology Department in a university located at east of China were selected for analysis. Those students were in the same department, so they read may read similar books, especially professional books. In addition, they interact frequently with each other as they attended classes together. We

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

intentionally chose all the first-year graduate students as participants as they formed a relatively close group. Then we analyzed the characters of this reading group to investigate social factors which may influence these students' reading.

The survey was sent out to all the first-year graduate students in Educational Department in a university in China. All students (42 students, 13 males and 29 females) responded and completed this survey. The data collected from the survey were analyzed using social network analysis. In this paper, we changed all participants' name with pseudonyms. In addition, six students were interviewed to for much more detailed information.

3 RESULTS

3.1 Major roles in the reading group

Figure 1 presents the book influence network (whom influence you in book selection) in this reading group. Based on our analysis, there were three different roles in group reading: outsiders, key leaders, and brokers. Long and Jie were outsiders in the reading group as they did not influence others' book choice. Apart from them, students, such as Yang, serverd as a leader in a reading group. In addition, there were also borkers in this reading group. In our analysis, there were multiple sub-groups. As shown in Figure 2, different groups are presented with different colors. To connect these sub-groups, brokers served as a bridge between those subgroups. For intance, without Yu, the red group cannot connect with the organe group as well as pink group. So, brokers play an important role in the information flow among different groups. Table 1 presents five types of brokers. All brokers played a role in connecting two or more groups.

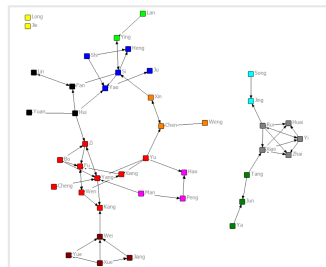


Figure 1: book influence network Figure 2: sub-groups in this class

Table 1: Five kinds of brokers

Types of brokers	Students
Coordinator	Yao, Chen, Peng, Yang, Wen, Xiang, Fu, Zi, Fan
Consultant	--
Gate keeper	Si, Yao, Chen, Yang, Zi, Fan, Xiao
Representative	Ying, Xin, Wei, Rui
Liaison	Ying, Si, Yao, Chen, Xin, Peng, Yang, Wen, Xiang, Fu, Zi, Fan, Wei, Yi, Rui, Xiao

3.2 Impact of reading behaviors

Correlation analysis was conducted to investigate how reading interactions (book borrowing, book attention, book sharing) influence group members' book choice. Table 2 provides the correlation results. The book borrowing behavior was the most influential factor, revealing that the flow of books impacts on book selection most in a reading group.

Table 2: Correlation analysis between influential relation and reading relations

	Book borrowing network	Book attention network	Book sharing network
Influential network	0.420***	0.386***	0.331***

*p<.05. **p<.01. ***p<.001

3.3 Impact of social behaviors

Table 3 shows the correlation analysis between book influence network and social relation network (emotion, informal communication, formal communication, advice, knowledge trust). Emotion network had the highest relation with book choice influential network. This revealed that readers tended to be influenced by people who they shared feelings with.

Table 3: Correlation analysis between influential relation and social relations

	Emotion network	Informal communication network	Formal communication network	Advice network	Knowledge trust network
Influential network	0.494***	0.475***	0.374***	0.408***	0.331***

*p<.05. **p<.01. ***p<.001

4 DISCUSSION AND CONCLUSION

The purpose of the study is to explore social features of a reading group and investigate social factors influencing people's book choice. In this paper, the author analyzed: 1) major roles in a reading group, 2) social factors (reading relations and social relations) impacting on group reading. The findings help to add valuable discussion to book choice in the era of social media. The significance of the study is to propose a new way to investigate social factors impacting people's book choice and provide insight for future research on book recommendation under social media environment.

REFERENCES

- Bang-Jensen, V. (2010). A children's choice program: Insights into book selection, social relationships, and reader identity. *Language Arts*, 87(3), 169-176.
- Dean, M. D. (2016). A call to embrace social reading in higher education. *Innovations in Education and Teaching International*, 53(3), 296-305.
- Despot, I., Ljevak Lebeda, I., & Tomašević, N. (2016). Social reading-the reader on digital margins. *Libellarium: časopis za povijest pisane riječi, knjige i baštinskih ustanova*, 9(1), 185-196.
- Vlieghe, J., Vandermeersche, G., & Soetaert, R. (2016). Social media in literacy education: Exploring social reading with pre-service teachers. *new media & society*, 18(5), 800-816.

The ACT Master(y) model for Measurement, Learning and Navigation

Gunter Maris, Steve Polyak, Michael Yudelson

ACTNext, by ACT

Gunter.maris@act.org

ABSTRACT: Historically, measurement and learning have evolved independently. More recently (e.g., Deonovic, et al., 2018) efforts have been made to close the gap and connect the two bodies of research together. Network models are at the center of this integration. To support an integrative real world data driven approach to integrated measurement, learning, and navigation of learners, models are needed that encode prior knowledge from both fields of research. The ACT master(y) model introduced here is such a model.

Keywords: network psychometrics, learning, measurement.

1 INTRODUCTION

The ACT Master(y) model has been developed at ACTNext to support **large scale** systems for personalized **learning, measurement** and **navigation**. Many models exist that support some of these, but no single model currently supports all of them. The Master(y) model builds on these partial solutions but is the only model that bridges the gap between assessment on the one hand and learning and navigation on the other hand. We introduce the key components below and explain briefly how they fit together.

2 KEY INGREDIENTS

Before introducing the ACT master(y) model proper, we lay out the key ingredients that together allow for baking the master(y) model.

2.1 Measurement

From over a Century of large scale assessment we've learned that assessment material (e.g., ACT test items) has a hierarchical structure. Mathematics items correlate more with other mathematics items than with reading items. Within mathematics, algebra items correlate more with other algebra items than with geometry items. This hierarchical structure is encoded in the **ACT Holistic Framework** (Camara et al., 2015), which is a key ingredient to the Master(y) model.

Every skill in the holistic (or any other) framework is conceptualized as being either mastered or not mastered. Every assessment item is tagged to a skill in the framework, with a different probability of a correct response for masters (a) and non-masters (b). With a mastery probability of p we obtain the following measurement model (where $a > b$) in Table 1.

Table 1: Basic measurement model

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

	Master	Non-Master
Correct	ap	$b(1-p)$
Wrong	$(1-a)p$	$(1-b)(1-p)$
Margin	p	$1-p$

However useful a framework with its associated measurement model is for assessment purposes; it does not quite help with deciding what a learner should do next.

2.2 Learning and Navigation

Building on foundational work by Jean-Claude Falmagne (implemented in his Aleks learning system), **knowledge graphs** have become one of the dominant models for learning and navigation. A knowledge graph encodes prerequisite relations between (fine grained) skills. It's no use teaching quadratic equations to learners that haven't mastered linear equations. These relations help in deciding whether a student is ready to learn a new skill, needs to study more on the skill she currently is working on, or revisit an earlier acquired skill which is preventing her from moving on.

2.3 Validity

Learners don't learn because they want to become skilled test takers, but to become skilled professionals (in their profession of choice). Almost any non-trivial activity in almost any profession requires using a (large) number of skills to get something done. The Master(y) model borrows ideas from **Cognitive Diagnosis Models** to link real life problems to skills, to help build the validity argument of a learning, assessment and navigation system to prepare learners for their profession of choice. Some skills, such as the cross cutting capabilities (aka, 21st century skills) in the holistic framework, are not easily assessed in isolation and out of context. The same structure allows for making inferences on (say) critical thinking, from items related to various cognitive skills.

3 THE ACT MASTER(Y) MODEL

The ACT Master(y) model combines these three ingredients in a single statistical model, which can support a scalable learning, measurement and navigation solution. Figure 1 below gives a graphical representation of the ACT Master(y) model. Squares denote observable variables (either item responses, or educational resources a learner has consumed), circles denote skills. Undirected edges encode the hierarchical structure of the Holistic Framework, and directed edges encode the prerequisite structure of a knowledge graph. The Q layer in the network serves to encode the dependence on multiple skills for observable variables.

4 THE ACT MASTER(Y) MODEL AT SCALE

A learning, measurement and navigation solution needs to support real-time skill tracking at scale. The ACTNext Recommendation and Diagnostics (RAD) Engine is an API that combines the power of an intelligent educational-content delivery platform with state of the art, real-time skill estimate tracking. Currently, the RAD API powers recommendations in [ACT Academy](#), but fully extends and integrates into any Learning and Assessment System (LAS), that aligns to any subject or set of

standards (i.e. Common Core, [ACT Holistic Framework](#), etc.) and *adaptively delivers relevant, free, and personalized content* to meet the needs of learners everywhere.

The combination of the ACT Master(y) model on the one hand and the RAD API on the other hand offers a flexible and scalable solution to support learning, measurement and navigation at scale.

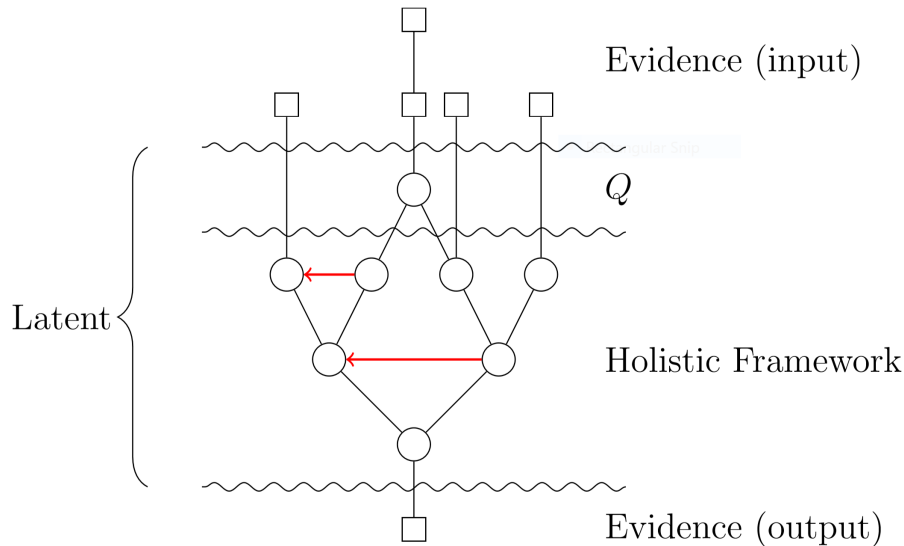


Figure 1: Graphical representation of the ACT Master Model

5 DISCUSSION

Over the last decade network models have gained wide popularity in psychology, and psychometrics in particular. A good overview of network psychometrics is found in Marsman, et al. (2018) and the book chapter by Epskamp, et al. (2018). Its impact goes from psychopathology, spear headed by Denny Borsboom (e.g., Borsboom & Cramer, 2013), to attitudes, spear headed by Han van der Maas (e.g., Dalege et al, 2016), and education(al measurement), spear headed by Gunter Maris (e.g., Savi et al., 2019).

The ACT master(y) model grew out of network psychometrics. It was developed to deal with Learning, Measurement and Navigation in an integrated fashion, but with interpretability, scalability, and applicability in real world learning systems in mind. To achieve these goals, we integrated as much prior (substantive) knowledge as we could into the modeling itself, thereby constraining it, but at the same time making it extremely tractable from a statistical point of view.

As always, when substantive knowledge meets data you can find out that the substantive knowledge was wrong or incomplete. Hence the ACT master(y) model is not an end point, but the start of a data driven iterative process of improvement to support the ultimate goals of Learning, Measurement and Navigation.

REFERENCES

- Borsboom, D., & Cramer, A. O. (2013). Network analysis: an integrative approach to the structure of psychopathology. *Annual review of clinical psychology*, 9, 91-121.
- Camara, W., O'Connor, R., Mattern, K., & Hanson, M. A. (2015). Beyond Academics: A Holistic Framework for Enhancing Education and Workplace Success. ACT Research Report Series. 2015 (4). ACT, Inc.
- Dalege, J., Borsboom, D., van Harreveld, F., van den Berg, H., Conner, M., & van der Maas, H. L. (2016). Toward a formalized account of attitudes: The Causal Attitude Network (CAN) model. *Psychological review*, 123(1), 2.
- Deonovic, B., Yudelson, M., Bolsinova, M., Attali, M., & Maris, G. (2018). Learning meets assessment. *Behaviormetrika*, 45(2), 457-474.
- Epskamp, S., Maris, G., Waldorp, L. J., & Borsboom, D. (2018). Network Psychometrics. In: Irwing, P., Hughes, D., & Booth, T. (Eds.), *The Wiley Handbook of Psychometric Testing, 2 Volume Set: A Multidisciplinary Reference on Survey, Scale and Test Development*. New York: Wiley.
- Marsman, M., Borsboom, D., Kruis, J., Epskamp, S., van Bork, R., Waldorp, L. J., ... & Maris, G. (2018). An introduction to network psychometrics: Relating Ising network models to item response theory models. *Multivariate behavioral research*, 53(1), 15-35.
- Savi, A. O., Marsman, M., van der Maas, H. L., & Maris, G. K. (2019). The wiring of intelligence. *Perspectives on Psychological Science*, 14(6), 1034-1061.

Evidence Based Decision Making in the Classroom

Geraldine Gray¹, Pauline Rooney¹, James Doody¹, Phelim Murnion², Kevin O'Rourke¹,
Lee O'Farrell³, Charles Lang⁴

¹Technological University Dublin,
Ireland
{Firstname.Surname}@tudublin.ie

²Galway-Mayo Institute of
Technology, Ireland
Phelim.Murnion@gmit.ie

³National Forum for the Enhancement
of Teaching and Learning, Ireland
Lee.OFarrell@teachingandlearning.ie

⁴Columbia University,
USA
cl3584@tc.columbia.edu

ABSTRACT: Although there is wide agreement on the benefits of Learning Analytics (LA), many institutes still struggle to operationalize their LA Strategy. While many stakeholders who generate data are enthusiastic about its potential value to improve the student experience, how to derive actionable intelligence from that data is perceived as a challenge. This workshop aims to explore initiatives and ideas that empower stakeholders in higher education to make better use of their data, to promote evidence-based practice and ultimately improve the student experience. Building on research currently being conducted in an Irish context, the workshop hopes to foster fresh thinking about LA that focuses on empowerment rather than surveillance, and ultimately promotes both self-efficacy and confidence among students and staff in their knowledge and use of big data in education.

Keywords: Engaging stakeholders, data literacy, operationalize learning analytics

1 WORKSHOP BACKGROUND

Faculty and professional services staff alike should be more involved in data-informed decision-making that impacts on student success. This connects the analysis of data with the learning or support environments that generated it. The SHEILA project¹ evidenced a strong interest in using learning analytics to enhance the student experience, particularly in areas such as the provision of timely feedback (Tsai et al. 2018). Student and staff focus groups conducted by this workshop's organizers concurred, and also confirmed SHEILA's findings of a preference amongst students to see learning analytics focus on improving the learning context rather than informing individual student interventions. Therefore, promoting and enabling more widespread use of learning analytics to enable data-informed teaching and learning practices is an important goal for the learning analytics community over the coming decade. One barrier to engaging with the relevant learning data is a lack of professional development and support in learning analytics for staff and students (Colvin, Dawson Wade and Gasevic, 2017), arguably augmented by poor digital skills generally (Bluck and Carter, 2016; DESI, 2018). Fostering ethical and effective use of student data to inform practice, and an ability to critique such usage with respect to individuals' rights, can serve as a driver for change.

¹ <https://sheilaproject.eu/>

Factors that impact on quality of learning vary from student to student, making students' digital footprint challenging to analyse and interpret correctly (Bergner, Gray, & Lang, 2018). However, correct interpretation of data is critical to maintaining quality in learning analytics and promoting evidence-based decision making (Gibons & Lang, 2018). While many data analytics tools are evolving to make the functionality of analysing and visualising data more user friendly, this does not promote critical analysis of the outputs. Staff and students have reported difficulty in interpreting learning data (Vieira, Parsons, & Byrd, 2018) and have also struggled to articulate what analytics would be useful for, as the possibilities were not well understood (Schumaer & Ifenthaler, 2016). As argued by Colvin et al. (2017), successful adoption of learning analytics must address a lack of data and digital literacy skills, a shortfall that is often overlooked by management.

Staff and students alike could benefit from resources focused on how to use student data for feedback on learning practices and, in the process, augment their own levels of digital and data literacy. Appropriate training and support can help students and staff recognize the potential and limitations of information derived from learning data, and thereby make better decisions from analysis of educational digital footprints. Ultimately, this will facilitate embedding data literacy into the curriculum in a way that will progress student intellectual development and critical thinking skills in a manner that is transferable to their life outside academia.

Underpinned by the theme of collaborative partnership between staff, student and sectoral communities, this workshop will focus on discussing the potential of strategies and training resources for staff and students in the use of learning data and the development of digital capability. We propose discussing how to bring learning analytics to the wider audience of practitioners in higher education, with the potential impact of more insightful analysis of learning data to support evidence-based practice. We aim to explore methods that can address skills and competency gaps that hinder staff and students from harnessing the potential of learning data.

2 ORGANISATIONAL DETAILS

The proposal is for a half day, open interactive workshop session covering introductions (15 mins), invited presentations on ongoing research that captures sector-wide initiatives relevant to the workshop theme, and a series of individual and group-based activities to promote discussion and sharing of ideas and experiences. Including breaks, the session will last 4 hours in total. A call to participate will be disseminated through relevant listservs and a workshop website. Participants are welcome to submit their areas of interest in advance of the workshop to facilitate birds of a feather discussion groups during workshop activities. The expected participant number is approximately twenty.

2.1 Required equipment

The standard equipment available in the room is sufficient along with wifi access and flipcharts.

3 WORKSHOP OBJECTIVES

Sharing of ideas and experiences

Evidenced-based learning analytics requires those in the cold face of teaching, learning and student support to harvest the information generated in their respective contexts, to inform best practice and optimise student success. Widespread acceptance and adoption of learning analytics to answer questions that have meaning for stakeholders is a key challenge over the coming decade. This workshop is a step towards exploring how that may be achieved.

Community Building

An objective of the workshop is to provide a meeting place for researchers who take a special interest in operationalizing learning analytics to support evidence-based teaching and learning practices. We anticipate that a concentrated meeting will promote continuing collaboration on this important topic.

Initiating a resource repository

Through community building and sharing ideals, a longer-term goal of this initiative is to establish a repository of resources aimed at improving data literacy in educational contexts.

4 THE WORKSHOP WEBSITE

The workshop's aims, objectives, organisers and activities are included on the work website², including a form to collate participants' interests in this area and related work done to date. Following the workshop, a summary of workshop submissions, discussions and key points will be added to the project website, subject to the permission of contributors. This will be disseminated through relevant listservs and twitter.

REFERENCES

- Bergner, Y., Gray, G., & Lang, C. (2018). What does methodology mean for learning analytics? *Journal of Learning Analytics*, 5(2), 1-8.
- Bluck, E. and Carter, A. (2016) *The e-Skills Manifesto (2016)*, European Schoolnet, Brussels, pp. 69 (<http://ec.europa.eu/DocsRoom/documents/21341/attachments/1/translations/en/renditions/pdf>)
- Colvin, C., Dawson, S., Wade, A., Gašević, D., Lang, C., Siemens, G., ... & Gašević, D. (2017). Addressing the challenges of institutional adoption. *Handbook of learning analytics*, 281-289.
- Conijn, R., Snijders, C., Kleingeld, A. and Matzat, U. (2017). Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS. *IEEE Transactions on Learning Technologies*, 10(1), 17-29.
- DESI (2018). *Digital Economy and Society Index (DESI) 2018: Country Report Ireland* (http://ec.europa.eu/information_society/newsroom/image/document/2018-20/ie-desi_2018-country-profile_eng_B4406C2F-97C3-AA9A-53C27B701589A4F3_52225.pdf)

² <https://daltai-he.ie/event/workshop-lak2020/>

- Gibson, A., & Lang, C. (2018, March). The pragmatic maxim as learning analytics research method. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 461-465). ACM.
- Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior*, 78, 397-407.
- Tsai, Y.-S., Gašević, D., Whitelock-Wainwright, A., Moreno-Marcos, P. M., Fernandez, A. R., Muñoz-Merino, P. J., Kloos, C. D., Tammets, K., Kollom, K., Scheffel, M., and Drachsler, H. (2018). *Teacher and Student Perspectives on Learning Analytics – Executive Summary*. Retrieved from https://sheilaproject.eu/wp-content/uploads/2018/06/Teacher-and-student-perspectives_3.pdf.
- Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education*, 122, 119-135.

DesignLAK20: Developing quality standards for analytic measures of learning for learning design

Sandra Milligan

University of Melbourne
s.milligan@unimelb.edu.au

Linda Corrin

Swinburne University of Technology
lcorrin@swin.edu.au

Nancy Law

University of Hong Kong
nlaw@hku.hk

Ulla Ringtved

University College of Northern Denmark
ulr@ucn.dk

ABSTRACT: The 5th Annual DesignLAK Workshop focuses on the challenge of ensuring quality measures of learning when using learning analytics in the context of learning design. The challenge of determining appropriate measures of learning forms part of a broader discussion in the learning analytics community around the methodological development of the field. When recognising that learning analytics practitioners encompass experts from diverse fields, who may not use a common set of methodologies, difficulties arise around how the quality of learning and design is assessed (Law et al., 2019; Ochoa, Herrchkovitz, Wise & Knight, 2017). In this interactive, half-day workshop participants will be given the opportunity to apply a newly developed quality standards framework to a range of learning design scenarios as a reflective and evaluative tool for design for learning. Participants will also have the opportunity to road-test the standards for their own context and to provide feedback to assist in further development and refinement of the standards framework. Outcomes from the workshop will include an improved version of the standards framework that will be made available to the learning analytics community and a research publication which will report on the evaluation and related discussions held during the workshop.

Keywords: Learning analytics, learning design, measures of learning, quality standards

1 BACKGROUND

When designing for learning (Goodyear, 2015) it is important to consider how such learning can be measured in order to evaluate whether the design meets its intended outcomes. Increasingly learning analytics is being used to facilitate such evaluation and consequently is dependent on measures of learning to provide meaningful and actionable information to stakeholders (Bergner, Gray & Lamb, 2018; Wilson & Scalise, 2016). Such measures can be used in a range of ways. Sometimes they are used to assess the degree of learning attained when judging the success of educational interventions. They may be used as targets in predictive or data mining studies. Or they

may be used as triggers in the development of instruments or apps which make judgements of learning outcomes or provide feedback on learning processes. The quality of the measures used – how valid, reliable and trustworthy they are – will shape the quality of the application of learning analytics to design (Bergner, 2017). Therefore, the use of learning analytics to evaluate and support learning design can be open to justified criticism if based on poor or doubtful measures of learning.

The challenge of determining appropriate measures of learning forms part of a broader discussion by the learning analytics community about methodological development within the field. When recognizing that learning analytics practitioners encompass experts from diverse fields, who may not use a common set of methodologies, difficulties arise around how the quality of learning and design is assessed (Law et al., 2019; Ochoa, Herrchkovitz, Wise & Knight, 2017). This is an important conversation to be had in the context of LAK2020 as we reflect on the development of the field over the past decade and identify ways to ensure quality and build impact into the future.

One way to address this challenge is to develop a common framework of standards to apply when using measures of learning as part of learning analytics processes and applications. These “metrolytic” (Milligan, 2018) standards can be used as reflective and evaluative tools when designing learning analytics for learning design, or when reviewing the work of others in the field. Application of metrolytical standards requires analysis of a range of indicators, addressing matters such as the clarity of the nature and scope of learning being measured; appropriateness of assessment methods used; the quality and character of the data being used as evidence; the quality of the argument as to the validity of the measures; the quality of evidence as to reliability and precision of measures; their suitability for purpose; the transparency of data transformation and algorithms applied; and the legitimacy of their interpretation from an educational perspective.

In the proposed workshop participants will be presented with a set of metrolytic standards to apply to different learning design contexts for the purposes of evaluation of quality and to highlight the importance of considering how learning can be effectively measured. The workshop is relevant to anyone involved in the design of learning and/or analytics who desires the ability to evaluate the quality of their work and measures. They will have the opportunity to road-test the standards for their own context and to provide feedback on the standards to assist in further development and refinement.

This workshop continues the conversation about the importance of the intersection between learning analytics and learning design which has been developed throughout the past four annual DesignLAK Workshops. In 2016 we considered how learning design could inform how learning analytics could be used to improve feedback practices (Milligan et al., 2016) and in 2017 examined the quality of learning analytics indicators for assessment design (Ringtved et al., 2017). The DesignLAK workshop in 2018 involved the evaluation of several systems and tools that linked learning analytics and learning design (Corrin et al., 2018) and in 2019 we examined how validity can be ensured in the use of learning analytics in assessment design (Law et al., 2019). The theme of this year’s workshop builds on the previous themes by bringing together the key issues raised around learning indicators, assessment, and design practices in an instrument (the metrolytic standards) that participants can use in their own context and which will be improved upon so it can form a contribution to the wider learning analytics community.

2 OBJECTIVE OF THE WORKSHOP

The main objective of the workshop is to explore the importance of identifying appropriate measures of learning for learning design and analytics through the evaluation of a set of newly developed metrolytic standards. The workshop will provide the authors of the standards an opportunity to evaluate their usefulness using a range of learning designs and contexts informed by the participants' own experiences as well as a set of pre-defined common learning designs used in higher education.

3 WORKSHOP DESIGN

This half-day workshop is designed to engage participants in the evaluation of the metrolytic standards within the context of learning design. The event will be interactive with opportunities for participants to work together in small groups and also to contribute to whole group discussions of key issues raised throughout the activities. Based on attendance at previous DesignLAK workshops we would expect approximately 20 participants. The workshop will open to anyone with an interest in learning analytics and learning design, both teachers/practitioners responsible for the implementation of learning designs/analytics and also researchers in the field.

3.1 Pre-workshop preparation

The workshop organisers will roll out a promotion strategy which will involve the advertisement of the workshop to a wide audience to attract a range of participants in roles related to learning analytics and learning design. To do this the organisers will utilise mailing lists of several professional societies associated with the field, as well as Twitter and other relevant social media. A website will be established for the workshop that will include information about the workshop and its design as well as a summary of the outcomes of the workshop post-event. Input on the learning design scenarios and initial feedback on the metrolytic standards will be sought from several leading learning design and learning analytics specialists, and their views will be captured to share with participants during the workshop.

3.2 The workshop

The workshop will open with a brief presentation on the topic of learning analytics and design with a focus on the importance of using appropriate measures of learning. Participants will then be given an opportunity to nominate any key questions that they have that they would like to see addressed throughout the workshop. This will be followed by an introduction to the metrolytic standards, including background on how they were developed and how it is envisaged that they can be applied in practice. A set of common learning design scenarios will then be given to the groups from which they can select one to evaluate using the metrolytic standards. Groups will be asked to work through the process in a systematic way, highlighting any issues or challenges and noting down any suggestions for improvement. At the end of this activity each group will be asked to report back on the evaluation they have conducted based on their assigned learning design and associated learning analytics. This will be followed by a whole group discussion of key issues raised. Participants will then be asked to form new groups which will be allocated based on their interest and/or experience of learning designs and analytics in their own context. The metrolytic standards will be applied to

their own context scenarios with opportunities given to provide feedback to the organisers on any key issues or suggestions for change that they foresee in their own practice. The workshop will end with a synthesis of all the issues and suggestions raised and a look at future development and contributions that the metrolytic standards could offer.

3.3 Dissemination of workshop outcomes

During the workshop participants will be encouraged to tweet about interesting concepts and discussions using the hashtag #DesignLAK20. A summary of the discussions will be curated by the workshop organisers in a Google Doc which will be shared with workshop participants who can also add further comments and questions. Participants will have ongoing access to the document after the event as a reference and as a venue to continue the conversation. During the workshop participants will be asked to complete a short survey about the utility of the metrolytical standards and asked to provide feedback that can inform the improvement of the standards as a practical tool for learning analytics practitioners. The outcomes of this workshop will form the basis of a research paper authored by the workshop coordinators for publication.

REFERENCES

- Bergner, Y., Gray, G., & Lang, C. (2018). What Does Methodology Mean for Learning Analytics?. *Journal of Learning Analytics*, 5(2), 1-8.
- Bergner, Y. (2017). Measurement and its uses in learning analytics. In C. Lang, G. Siemens, A. F. Wise, & D. Gašević (Eds.), *Handbook of Learning Analytics* (pp. 34–48). Alberta, Canada: Society for Learning Analytics Research (SoLAR). <http://dx.doi.org/10.18608/hla17.003>
- Corrin, L., Law, N., Milligan, S., & Ringtved, U. L. (2018). Workshop: Methodological Bases for the Measurement of Learning in Learning Analytics. In *Companion Proceedings of the 8th International Conference on Learning Analytics and Knowledge*. ACM.
- Goodyear, P. (2015). Teaching as design. *HERDSA Review of Higher Education*, 2, 27-50.
- Law, N. W. Y., Corrin, L., Milligan, S., & Ringtved, U. (2019). DesignLAK19: Ensuring validity in assessment design and analytics. In *Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge*. ACM.
- Milligan, S. K. (2018). Methodological foundations for the measurement of learning in learning analytics. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 466-470). ACM.
- Milligan, S., Ringtved, U., & Corrin, L. (2016). Learning design and feedback processes at scale: stocktaking emergent theory and practice. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 479-480). ACM.
- Ochoa, X., HersHKovitz, A., Wise, A., & Knight, S. (2017). Towards a convergent development of learning analytics. *Journal of Learning Analytics*, 4(3), 1-6.
- Ringtved, U., Milligan, S., Corrin, L., Littlejohn, A., & Law, N. (2017). Quality Metrics and Indicators for Analytics of Assessment Design at Scale. In *Proceedings of the Seventh International Conference on Learning Analytics & Knowledge* (pp. 508-509). ACM.
- Wilson, M., & Scalise, K. (2016). Learning analytics: Negotiating the intersection of measurement technology and information technology. In M. J. Spector, B. B. Lockee, & M. D. Childress (Eds.), *Learning, design, and technology* (pp. 1–23). Springer, Cham. http://dx.doi.org/10.1007/978-3-319-17727-4_44-1.

Building Capacity Through the Learning Analytics Learning Network

Justin T. Dellinger

University of Texas at Arlington
jdelling@uta.edu

Florence Gabriel

University of South Australia
florence.gabriel@unisa.edu.au

Ryan Baker

University of Pennsylvania
rybaker@upenn.edu

Shane Dawson

University of South Australia
shane.dawson@unisa.edu.au

George Siemens

University of Texas at Arlington, University of South Australia
gsiemens@uta.edu

ABSTRACT: In spite of the rapid growth of data science, the development of learning scientists and analysts is often challenging given the lack of formal degree programs. An efficient approach to developing the skills needed by professionals to engage in learning analytics is to leverage existing communities of practice and networks. To address this issue, we developed the international Learning Analytics Learning Network (LALN). LALN events are held monthly, and speakers from 25 nodes (a local or regional learning analytics community) across the world take turns to introduce participants to new methods and learning activities. This workshop will start with presentations by Professor George Siemens and Professor Shane Dawson. Participants will then explore how to best capitalize on this network and discuss the needs of their own community. Finally, a panel of LALN experts will discuss the challenges and opportunities of this network.

Keywords: Learning Analytics; International Network; Learning Network; Data Science; Community of Practice; Networks of Practice

1 INTRODUCTION TO THE LEARNING ANALYTICS LEARNING NETWORK

Data science has emerged as an important part of educational research and practice in recent years, producing a rapidly growing demand for a workforce that is literate in data science methods as well as competence in the specific nature of educational data, research, and practice (Baker & Siemens, 2014). However, there is not a sufficient number of graduate programs or other sustained training activities to meet this need. As a result, much of the learning analytics workforce lacks key competencies.

To address this gap, a consortium consisting of scholars from the University of South Australia, the University of Texas at Arlington, and the University of Pennsylvania have developed an international Learning Analytics Learning Network (LALN). Monthly meetups are being held worldwide – local research community leaders in 25 cities have agreed to participate, from New York City and Silicon Valley to Kyoto, Manila, and Frankfurt.

Cities take turns hosting a distinguished speaker, streaming the event online so other cities can join. Events are also recorded for later viewing to accommodate the different time zones. Local moderated discussions are then held. Activities and exercises are focused on beginner, intermediate, and expert categories. They range from introducing participants to learning analytics to helping them learn to use modern and emerging cyberinfrastructure for data science (including activities such as Python and R in cloud computing) and deploying common learning analytics algorithms (such as Bayesian Knowledge Tracing) efficiently at scale through cloud infrastructure. Our activities serve both as an introduction to methods for new members in the field (such as graduate students and teachers) and as continuing education for existing members of the research workforce, responsive to changes in the tools, algorithms, and the technologies needed for data science.

Traditional approaches to building capacity are slow to scale or limited in scope (Dawson et al., 2019). The aim of the LALN is to develop a network of practice, where regional communities come together to collaboratively create resources and learning experiences, leveraging network effects where each additional member increases the benefits and usefulness of being part of the network. We use formative feedback and data to improve our network of practice, making it sustainable long-term even as it scales to more cities around the globe. Learning analytics is underpinning the emergence of key advances in education such as adaptive learning, at-risk prediction, and intervention; our network will speed up the deployment of existing technologies as well as the development of new technologies that will increase student achievement.

2 OBJECTIVES

The objectives are this workshop are to:

- *Introduce the LALN to the LAK community.* LALN is a global network for networked professional development. The first events were held in October 2019. While there is some overlap with the existing LAK community, there are numerous regional nodes that are not affiliated with LAK. We expect that many LAK researchers will find value in engaging with LALN for their professional development and also for their students to join a global analytics community.
- *Explore how LALN can best address the needs of the LAK community.* LALN is a grassroots distributed network, connecting regional communities. Research topics and tutorials are locally organized. This approach allows ideas to spread bottom-up rather than in a planned top-down approach. As such, a key objective of the workshop is to hear from the LAK community regarding the types of organizational strategies, frequency of events, and related networking activities that they would find valuable.

3 PROPOSED ACTIVITIES

3.1 Presentations

This workshop will start with presentations, requested and confirmed, by George Siemens, Justin T. Dellinger, and Shane Dawson. Dawson will detail the challenges of moving research from small lab environments to broader scale adoption. Siemens and Dellinger will address the structure and operation of LALN. These presentations will set the context for group brainstorming and discussions that will follow.

3.2 Brainstorming session

Participants will then be invited to a brainstorming session where we will be able to explore and detail the needs that LALN could support. We will discuss topics such as organizational strategies, the frequency of events and networking activities that LAK participants would benefit from. This session will start with participants working in small groups (3 to 5 people) and will conclude with a whole group discussion.

3.3 Panel discussion

We have reached out to academics who are running local learning analytics communities. These experts will examine the opportunities and challenges of LALN. They will also discuss how to build communities and run sessions based on their experience.

4 WORKSHOP FORMAT

We proposed a half-day open workshop and expect up to 40 participants. A website has been created to share all relevant resources generated before and after the workshop (e.g., key readings, presentations, video clips, discussion notes, and documentation for joining LALN) as is available at <https://sites.google.com/view/laln/events/lak20-workshop>. For this workshop, we will need a conference room with a capacity for up to 50 people with a setup that allows for small group discussions. A computer, screen, and Apple adapter will be needed for presentations, as well as A3 sheets, markers, and a flipchart to summarize the brainstorming session.

REFERENCES

- Baker, R., & Siemens, G. (2014). Educational data mining and learning analytics. In K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (2nd ed., pp. 253-274).
- Dawson, S., Joksimovic, S., Poquet, O., & Siemens, G. (2019, March). Increasing the impact of learning analytics. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 446-455). ACM.

Learning Analytics Principles of Use: Making Ethics Actionable

Kimberly E. Arnold

University of Wisconsin-Madison
kimberly.arnold@wisc.edu

Marcia Ham

The Ohio State University
Ham.73@osu.edu

Robin Pappas

Oregon State University
Robin.Pappas@oregonstate.edu

George Rehrey

Indiana University
grehrey@indiana.edu

ABSTRACT: This half-day interactive workshop, open to anyone, is designed to continue community discussions around making ethics actionable for learning analytics practitioners. Using a strong theoretical underpinning, the workshop focuses on how to advance the practical application of learning analytics principles / codes of practice at various institutions. Participatory discussions form the basis of this workshop where participants engage with resources and peers to develop plans for action to take back to their institution.

Keywords: learning analytics, ethics, privacy, code of practice, principles of application

1 WORKSHOP BACKGROUND

Learning Analytics (LA) as an educational practice is undertaken as a way to optimize the teaching and learning environment. As infrastructure, policies, and culture in higher education shift to support the need, educators have the opportunity to begin leveraging extensive amounts of data to support teaching and learning. The excitement surrounding all the possibilities that a data-guided culture offers to higher education is understandable. This includes the promise of predictive analytics, innovative new teaching practices, improved student success, and institutional transformation. However, as we embrace the application of LA, it is imperative that we do not overlook the ethical dimensions of these new practices, as well as unanticipated future developments. This is especially relevant in a time when: data breaches are becoming more prevalent in education^{1,2}; there is a heightened concern about an oncoming surveillance culture

¹ <https://www.idtheftcenter.org/at-mid-year-u-s-data-breaches-increase-at-record-pace/>

² <https://www.statista.com/statistics/273550/data-breaches-recorded-in-the-united-states-by-number-of-breaches-and-records-exposed/>

(Lyon, 2017; Shaw, 2017); and institutions are contracting with private vendors to improve student success. Evidence of this neglect and a sign of how pervasive this concern has risen, can be found in a recent letter sent by U.S. Congress to large EdTech companies and “data brokers”³. Congress demands more accountability and transparency surrounding the use of student data.

A search of scholarly literature shows increasing attention is being paid to the issue of ethics in LA (Drachsler & Greller, 2016; Ifenthaler & Tracey, 2016; Rubel & Jones, 2016; West, Huijser, & Heath, 2016; Pardo & Siemens, 2014; Slade & Prinsloo, 2013). However, applied ethics have not become pervasive in the sphere of LA practice. Making ethics actionable for institutions engaged in the use of LA offers unique challenges. Rarely are the people who make practical use of the tools and data systems the same ones with responsibilities for creating and implementing compliance policies, although a few institutions have found ways to address this situation (Berman, S., Daniel, M., Ham, M., & Robinson, P., 2018). This can lead to LA principles and codes of practice being crafted in a theoretical vacuum, far from the practicalities of implementation. For example, the right of erasure clause of the European General Data Protection Regulation (GDPR) states that individuals have the right to be forgotten⁴. While relevant organizations (e.g. AACRAO⁵ and NSC⁶) are forming conversations about GDPR, to our knowledge, very few enterprise level educational technologies have any function that allows the application of that clause to be enacted.

This workshop focuses on the intersection of ethics theory and the application of learning analytics. By leveraging LA-specific principles and codes or practice derived from theory and policy, it is possible to make the application of ethics accessible to all, leading to more inclusive, proactive conversations. This workshop seeks to generate and energize a community of like-minded individuals to take action in this realm back to their local institution and commit to iterative work in this fluid space. This workshop is designed to develop practical solutions for establishing and operationalizing learning analytics guiding principles and codes of practice.

1.1 Motivation for the Workshop and Relevance to the Field

Learning analytics practitioners have few resources at their disposal for applying realistic ethical frameworks. LA does not have a commonly accepted professional code of ethics or set of guiding principles for practice, in part due to the exceptionally wide application of educational practice. National laws such as FERPA provide some guidance, but interpretation varies across institutions. Further, existing, privacy policies are not sufficient to fully cover LA, so there is a very real need to move beyond the status quo and into a reality where learning analytics specific policies exist.

³ <https://www.durbin.senate.gov/imo/media/doc/Google-Pichai.pdf>

⁴ <https://gdpr-info.eu/issues/right-to-be-forgotten/>

⁵ <https://www.aacrao.org/resources/newsletters-blogs/aacrao-transcript/transcript/aacrao-releases-gdpr-guide>

⁶ <https://studentclearinghouse.org/nsblog/aacrao-and-clearinghouse-leaders-meet-to-discuss-technology-compliance-reporting-veterans-comprehensive-learner-record-and-ferpa/?hilite=%27gdpr%27>

1.2 Relevance to Conference Attendees, LAK Research Community, and the Field

This workshop is directly applicable to the LAK2020 theme of “shaping the future of the field” because of its focus on making ethics accessible to everyone, particularly practitioners and researchers. Having meaningful and proactive conversations about the ethics of learning analytics will directly translate into a higher likelihood of trust in the process of learning analytics as well as adoptions of tools and methods. Without pervasive conversations of this nature, long term sustainability of learning analytics may not be guaranteed.

2 WORKSHOP DETAILS

This workshop, open to anyone in the community, will provide a high-level landscape on what policies/practices/ethics statements exist around ethics and data privacy in learning analytics, as well as introduce a community-built resource library. However, the majority of the workshop will focus on participatory discussions about the importance of proactively incorporating ethics and student privacy in LA initiatives. The intent is to provide an open space for discussion of a complex issue, which manifests itself differently in different courses, institutions, states and countries.

2.1 Workshop Organizers

All workshop organizers are from institutions that have made significant investment in trying to scale and sustain LA, including learning how to move from research to practice in culturally appropriate ways. Ethical discourse and student privacy have been central tenants, and each institution has leveraged community to find pragmatic approaches to principles, heuristics, and policies surrounding learning analytics. Making ethics “real” to various stakeholder groups has benefited each institution in different ways. The facilitators are housed in different units at their institution: academic positions, administration, teaching excellence centers, and academic technology units, demonstrating that diverse stakeholders and perspectives are crucial to success. The breadth of experience and methods used to accomplish this across five institutions will be brought to bear in this workshop.

2.2 Workshop Activities

The interactive workshop will be discussion based and outcome driven. Participants will have access to a curated set of resources related to ethics, policy, and privacy in learning analytics, as well as a library of tools that can be applied in their local context for building a narrative. Major activities of the workshop will focus on the creation and application of learning analytics ethical principles / code of conduct. Discussions consider the importance of showing the benefit of leveraging principles/code of conduct (such as Sclater’s, 2016) in decision making processes and how generic processes can be created for the greater good. Participants will also work through contextualizing action plans for their institution based on personal reflection and group feedback. A focus will also be placed on how being part of a community helps shape the way we think about these issues.

2.3 Workshop Outcomes and Dissemination

This workshop will have many pragmatic outcomes. First, each participant will leave with an action plan they create that is contextualized for their institution and that will be the first step in creating a

roadmap for adopting/adapting principles or codes of practice to help bridge theory and practice. These plans will be drafted and revised based on workshop feedback. Mechanisms will also be put in place for ongoing conversation, iterations, and knowledge sharing after the workshop. Secondly, the workshop facilitators will synthesize the conversations and the collaboratively generated artifacts and create a white paper with the outcomes of the workshop. Finally, the workshop will serve as an additional point in the continued solidification of a community of transformation in which the application of ethical principles for learning analytics practice is central.

These outcomes are ambitious, but the facilitators are dedicated to carrying the work beyond a single, day-long workshop. Sustained engagement will occur through an ongoing community of practice with a focus on making ethics accessible and actionable to various stakeholder groups. The white paper produced following this workshop will be disseminated within the SoLAR community as well as through learning analytics practitioner communities such as EDUCAUSE, The Learning Analytics Community Exchange, the Unizin Consortium, and other institutional and consortial communities of practice/inquiry.

REFERENCES

- Berman, S., Daniel, M., Ham, M., & Robinson, P. (2018, October). *Learning analytics: Integrating Privacy and Ethics for Improved Student Outcomes*. EDUCAUSE Annual Conference, Denver.
- Drachler, H., & Greller, W. (2016, April). *Privacy and analytics: it's a DELICATE issue a checklist for trusted learning analytics*. Sixth International Conference on Learning Analytics & Knowledge. Edinburgh, Scotland, 2016.
- Ifenthaler, D., & Tracey, M. W. (2016). Exploring the relationship of ethics and privacy in learning analytics and design: implications for the field of educational technology. *Educational Technology Research and Development*, 64(5), 877-880.
- Lyon, D. (2017). Digital citizenship and surveillance | surveillance culture: Engagement, exposure, and ethics in digital modernity. *International Journal of Communication*, 11, 19.
- Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438-450.
- Rubel, A., & Jones, K. M. (2016). Student privacy in learning analytics: An information ethics perspective. *The Information Society*, 32(2), 143-159.
- Sclater, N. (2016). Developing a Code of Practice for Learning Analytics. *Journal of Learning Analytics*, 3(1), 16-42.
- Shaw, J. (2017). The Watchers. Assaults on privacy in America. *Harvard Magazine*. Retrieved from <https://harvardmagazine.com/2017/01/the-watchers>
- Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510-1529.
- West, D., Huijser, H., & Heath, D. (2016). Putting an ethical lens on learning analytics. *Educational Technology Research and Development*, 64(5), 903-922.

3rd Personalising feedback at scale Workshop: Teacher-driven action and dialogue

Vigentini, Lorenzo

UNSW Sydney

l.vigentini@unsw.edu.au

Sarah, Howard

U. of Wollongong

sahoward@uow.edu.au

Liu, Danny Y.T.

University of Sydney

danny.liu@sydney.edu.au

Lim, Lisa

U. of South Australia

lisa.lim@unisa.edu.au

ABSTRACT: After two successful workshops at LAK, in which presenters explored tools used to provide feedback at scale (LAK18) and focused on students (LAK19), this workshop shifts the attention to teachers and how they can use data-driven approaches to support the provision of feedback. Stories ‘from the trenches’ with a focus on educators and practitioners is what will drive the workshop. Continuing to bring together scholars and practitioners to find a common ground for showcasing interesting examples of effective feedback, the workshop will showcase what and how data can be used to improve the process and richness of feedback for both learners and educators. Key outcomes will be a better understanding of approaches and existing cases of good practice which will foster discussion and collaboration in the LA community.

Keywords: personalization, effective feedback, student-centred analytics

1 INTRODUCTION

1.1 Background

With two successful workshops delivered at LAK which explored the issues of tools developed to support the scaling of personalized feedback and how these benefits students, this third workshop shifts the attention to educators and practitioners in the feedback process.

The provision of effective and timely feedback *of* and *for* learning (Brown & Knight, 1994; Hattie, 2008; Hattie & Timperley, 2007; Hounsell, 2003) is essential in influencing students’ achievement, promoting autonomy and self-regulation (Nicol, 2010; Sadler, 2010). However, feedback and assessment are often the lowest rated aspect in terms of satisfaction from students’ evaluations and satisfaction surveys (Radloff, Coates, James, & Krause, 2011; Williams & Kane, 2008).

Generally, there are three reasons why feedback fails: 1) students do not perceive feedback as feedback, or they don’t use it effectively; i.e. there are differences in how students and teachers understand feedback (Pitt & Norton, 2017; Henderson, Ryan, & Phillips, 2019; O’Donovan, Price, & Lloyd, 2019); 2) the feedback is not appropriate or good enough (this could be because of how feedback is provided or because the focus and purposes are not shared: i.e. providing feedback is reduced to a summative, corrective and transmissive process; see Carless, 2006; Forsythe & Johnson, 2017); 3) teachers fail to

provide added value (i.e. too generic or superficial, providing a *final* judgement on students' submitted assignments instead of an opportunity to reflect and improve (see Nicol, 2010; Carless, et al, 2011; Carless, 2019). Further, the quality and value of feedback is worsened by the increased workload, rise in the numbers of students and the shrinking of teaching periods (Pace et al, 2019), which makes it harder for teachers to focus on individual assessments and/ or provision of personalised, and actionable feedback (Evans, 2013; Carless, 2019).

Some educational researchers have started to reconsider the impact (or lack) of feedback as currently implemented in Higher Education (Forsythe & Johnson, 2017; Pitt & Norton, 2017) and focus more on the constructive value of a dialogic approach in which both giving and receiving feedback are considered more holistically (Forsythe & Johnson, 2017; Nicol, 2010; Pitt & Norton, 2017). This is more akin to the model of continuous feedback which students are used to in schools (Hattie, 2008), but in order to implement it effectively, teachers require access to granular performance over time and to use this to engage with students frequently enough to build effective learning. Moreover, contemporary definitions of feedback emphasise that it is "a process through which learners make sense of information from various sources and use it to enhance their work or learning strategies. This perspective places emphasis on student engagement with feedback in terms of the shorter-term, e.g. improving performance on a piece of work, or longer-term, e.g. improving strategies for approaching assessment tasks. When information leads to actions, a feedback loop is said to be closed." (Carless, 2019:706). In light of this, a key challenge for LA-enabled feedback is a move away from just 'feedback as telling' and enabling feedback to close the loop for students' learning.

While scholars and practitioners in the learning analytics (LA) field have made tangible connections with critical aspects that can shape learning, such as learning design and self-regulation, the provision of feedback to students has been relatively neglected (Liu et al., 2017; Pardo, 2017). This is despite the affordances of LA tools to understand and improve learning by "informing and empowering instructors and learners" (Siemens & Baker, 2012). Interestingly, despite a broad consensus that LA has the potential to positively affect education (Ferguson et al., 2016; Siemens & Baker, 2012), its adoption in Higher Education Institutions (HEI) has been sluggish (Ferguson & Clow, 2017), and "the pace of adoption of analytics within education organizations can be categorized as at best sporadic, and at worst resistant" (Dawson et al., 2018, p. 237). Several obstacles have been described, including issues around data (quality, access, ownership, data literacy, analytical capacity), organisational landscape (governance, culture, preparedness, funding) and technical aspects (implementation, adoption, change-management, availability of the 'right' tools (Bichsel, 2012; Ferguson et al., 2016; Tsai & Gasevic, 2017). Newer LA systems are starting to support teachers with means to provide rich feedback beyond typical early warning messages (e.g. SRES or Ontask - Liu et al., 2017; Pardo et al 2018; Tempelaar, Rienties, & Giesbers, 2015), but it is clear that there is a need and appetite in the LA community for research and practice to further explore data-informed student-centered pedagogies to provide feedback at scale.

Teachers need concrete approaches and support mechanisms to bridge the gap between LA research and classroom practice. This third workshop at LAK specifically focuses on educators and practitioners, their experiences and their stories in engaging with feedback and assessment practices supported by LA tools and approaches.

2 SCOPE OF THE WORKSHOP

This workshop brings together scholars and practitioners to explore examples of how educators use information (data) to enhance the feedback process for increasing students' engagement and

performance by scaffolding their learning processes with appropriate feedback on both content and strategies. The workshop has three primary goals:

1. Give a multidisciplinary theoretical foundation for practitioners and researchers in LA for effective data-informed feedback practices in HE;
2. Showcase extant or planned approaches for scaling feedback and consider how students receive and use the feedback; special attention is on approaches that are data-driven and personalised;
3. Promote reflection on both pedagogical and technological approaches to improve feedback practices targeted at the improvement of student learning and their ability to self-regulate learning.

3 ORGANISATION DETAILS

This will be a half-day workshop with mixed participation (including selected presenters and interested delegates). The organisers will welcome short or 'in progress' papers covering a range of issues including:

- 1) Overview of tool(s)/approach(es) to personalise feedback;
- 2) Implementation process (e.g. infrastructural, staff capacity, etc.);
- 3) Challenges and successes (as well as failures);
- 4) Stakeholder engagement, buy-in, and impact (especially faculty, students).

3.1 Who is this workshop for?

Those who wish to understand and apply principles of 'good' data-driven feedback for learning are welcome (including Educators/teachers and researchers, Technologists and educational developers, Learning scientists and data scientists/analysts and Academic managers). Given the explicit multidisciplinary nature of the workshop we expect that it will provide an opportunity to discuss and share innovations, impact on learning, and explore future directions in the application of learning analytics (LA) to personalisation of feedback.

3.2 Proposed workshop activities

After a brief introduction and conceptualisation of the workshop, a series of short presentations will provide a backdrop and provocation to think about current feedback practices: this will consider both the typically sparse provision in Higher Education as well as the continuous provision typical of schools. We will discuss both successful and unsuccessful approaches to better understand what works, in which context and for what type of students. Ample opportunity for discussion will be provided to address key themes and issues surfaced during presentations. Similar to the previous two Workshops, a website will be created to provide access to all contributions and presentations as well as a summary from the organisers after the workshop. The workshop will provide an avenue to continue the conversations beyond the session and open opportunities for further collaborations.

4 INTENDED OUTCOMES FOR PARTICIPANTS

We expect a range of presentations that will cover practical, evidence-based approaches to personalising data-driven feedback at scale with a focus on teacher and practitioner perspectives. We expect that participants will obtain a broad perspective of different approaches to using data for personalising feedback. They will enhance their understanding of the forms of feedback that could improve student learning by discussing cases, issues, and potential solutions to implementing LA-enhanced feedback practices. They will also have an opportunity to connect with researchers and practitioners working to provide personalised feedback, yielding opportunities for collaboration on approaches and tools across attending institutions. After the workshop, given the commitment to further collaborations, contributors will be invited to consider more substantial submissions with the intention to collate the works into a special issue of journal or an edited book on the topic.

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

5 REFERENCES

- Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education*, 38(6), 698–712. doi:10.1080/02602938.2012.691462
- Brown, S., & Knight, P. (1994). *Assessing learners in higher education*. Psychology Press.
- Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, 31(2), 219–233. <https://doi.org/10.1080/03075070600572132>
- Carless, D. (2019). Feedback loops and the longer-term: Towards feedback spirals. *Assessment & Evaluation in Higher Education*, 44(5), 705–714. <https://doi.org/10/gf33kc>
- Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education*, 36(4), 395–407. <https://doi.org/10/ftvjq4>
- Evans, C. (2013). Making Sense of Assessment Feedback in Higher Education. *Review of Educational Research*, 83(1), 70–120. <https://doi.org/10/gf82tm>
- Forsythe, A., & Johnson, S. (2017). Thanks, but no-thanks for the feedback. *Assessment & Evaluation in Higher Education*, 42(6), 850–859. <https://doi.org/10.1080/02602938.2016.1202190>
- Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. routledge.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Henderson, M., Ryan, T., & Phillips, M. (2019). The challenges of feedback in higher education. *Assessment & Evaluation in Higher Education*, 1-16. doi:10.1080/02602938.2019.1599815
- Hounsell, D. (2003). Student feedback, learning and development. *Higher Education and the Lifecourse*, 67–78.
- Liu, D. Y.-T., Bartimote-Aufflick, K., Pardo, A., & Bridgeman, A. J. (2017). Data-Driven Personalization of Student Learning Support in Higher Education. In *Learning Analytics: Fundamentals, Applications, and Trends* (pp. 143–169). Springer, Cham. https://doi.org/10.1007/978-3-319-52977-6_5
- Nicol, D. (2010). From monologue to dialogue: improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, 35(5), 501–517. <https://doi.org/10.1080/02602931003786559>
- O'Donovan, B. M., den Outer, B., Price, M., & Lloyd, A. (2019). What makes good feedback good? *Studies in Higher Education*, 1-12. doi:10.1080/03075079.2019.1630812
- Orsmond, P., Maw, S. J., Park, J. R., Gomez, S., & Crook, A. C. (2013). Moving feedback forward: theory to practice. *Assessment & Evaluation in Higher Education*, 38(2), 240–252. <https://doi.org/10.1080/02602938.2011.625472>
- Pace, F., D'Urso, G., Zappulla, C., & Pace, U. (2019). The relation between workload and personal well-being among university professors. *Current Psychology*. <https://doi.org/10/gf3kqp>
- Pardo, A., Bartimote-Aufflick, K., Buckingham Shum, S., Dawson, S., Gao, J., Gašević, D., . . . Vigentini, L. (2018). OnTask: Delivering Data-Informed Personalized Learning Support Actions. *Journal of Learning Analytics*, 5(3), 235–249. doi:10.18608/jla.2018.53.15
- Pitt, E., & Norton, L. (2017). 'Now that's the feedback I want!' Students' reactions to feedback on graded work and what they do with it. *Assessment & Evaluation in Higher Education*, 42(4), 499–516. <https://doi.org/10.1080/02602938.2016.1142500>
- Radloff, A., Coates, H., James, R., & Krause, K.-L. (2011). Report on the Development of the University Experience Survey. *Higher Education Research*. Retrieved from http://works.bepress.com/hamish_coates/79
- Sadler, D. R. (2010). Beyond feedback: developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35(5), 535–550. <https://doi.org/10.1080/02602930903541015>
- Siemens, G., & Baker, R. S. J. d. (2012). Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. In *Proceedings of the 2Nd International Conference on Learning Analytics and Knowledge* (pp. 252–254). New York, NY, USA: ACM. <https://doi.org/10.1145/2330601.2330661>
- Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior*, 47(Supplement C), 157–167. <https://doi.org/10.1016/j.chb.2014.05.038>
- Vigentini, L., Liu, D. Y.-T., Lim, L., & Martinez-Maldonado, R. (2018). Personalising feedback at scale: approaches and practicalities. In *Companion Proceedings of the 8th International Conference on Learning Analytics & Knowledge (LAK'18)*. Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education*, 31(3), 379–394. <https://doi.org/10.1080/02602930500353061>
- Williams, J., & Kane, D. (2008). Exploring the National Student Survey Assessment and feedback issues. *The Higher Education Academy, Centre for Research into Quality*.

Analytics-based Personalised Feedback and Learning Outcomes in ESL at School Level

Deepti Yadav

Regional Institute of Education, NCERT,
Bhopal, India
drbs.contact@gmail.com

N. C. Ojha

Regional Institute of Education, NCERT,
Bhopal, India
ncojhanerie@gmail.com

ABSTRACT: Though Learning Analytics (LA) is conceived as a substantial tool in higher education, according to a study by Viberg et al. (2018), it still needs more empirical studies and further research, especially in early and school education and thereby improving learning outcomes. The presentation focuses on the usage of LA in school education. The aim of the work was to explore Moodle and Google platforms Learning analytics capabilities to improve learning outcomes in English as Second Language (ESL). The objective was to study the learning progression and achievement in ESL's Basic Interpersonal Communication Skills (BICS), Cognitive Academic Language Proficiency (CALP) and overall scores of the learners in the successive trials of Class XI learners when Analytics-based Personalised Feedback is provided on the basis of performance with the help of Google forms. With Google forms, analytics was used to map the progression and performance of the learners by the teacher. Time-series design for a single group was used for the study. The data were collected over a period of three months from the thirty-nine students selected randomly in Bhopal city of India who enrolled in the [My Learning Class](#) (MLC) course using the Moodle platform, and accordingly personalised feedback and improvements were provided. Data were analyzed using Repeated Measures ANOVA. Results, findings and educational implications of the study have been discussed.

Keywords: Learning Analytics, BICS and CALP, ESL, Moodle, and Google

1 INTRODUCTION

Information and communication technology (ICT) in the field of education for pedagogical usage advances in the four stages of emerging, application, infusion and transformation. These stages differ according to the technological advances, development and growth of a country. In India, ICT is undergoing a transformational change from stage three to stage four with its online courses and modules. Indian classrooms are multilingual and multi-cultural. The policy-makers have tried to adopt three-language formula and English is taught as second language in India. Cummins (1981) modelled Basic Interpersonal Communication Skills (BICS) and Cognitive Academic Language Proficiency (CALP). Accordingly, there is a threshold level in a learners' first language that learners should achieve in CALP before they can achieve academic success in the second language acquisition. BICS refers to conversational fluency in a language and CALP refers to learners' ability to understand and express academic concepts and ideas in oral and written modes. The study delves on how teachers can implicate tools of LA in teaching-learning of these in-built components with the help of personalised feedbacks. The results of the work will provide the base for further deliberations and discussions.

Analytics is applied worldwide in all the fields for discovery, interpretation, and communication of significant patterns in information collected. These patterns are further used and applied for effective decision making. Analytics in education is extensively used for improving learning, feedbacks, student-

retention, administrative purposes and related decision making. For the purpose of feedback and language teaching and learning, LA has been explored theoretically and practically in several studies. Pardo et al. (2019) presented an approach which was associated with a positive impact on students' perception of quality of feedback and further on academic achievement through their case study; Admiraal and Bulterman-Bos (2017) carried out a case study with five secondary Dutch language teachers using online performance data of their students; and Peng (2017) tried to explore English teaching and learning modes based on learning analytics. All these studies provide further scope of empirical researches in the role of feedback and ESL with the two components of BICS and CALP.

2 EXPLORING FEEDBACK TO ENHANCE LEARNING OUTCOMES

For the study, LA has been contextualized as to provide Analytics-based Personalised Feedback (AbPF) to learners after observing performance through their Google forms submitted as content responses to improve learning outcomes. BICS and CALP are two in-built components of ESL teaching at Higher Secondary level in Indian schools. The study explored two learning platforms Moodle (version 3.6.1) LMS and Google in Indian context being cost-effective. The focal purpose of the study was to provide AbPF in ESL components to ultimately improve learning outcomes in terms of achievement in BICS, CALP and Overall score of learners of Class XI.

2.1 Hypothesis Explored in the Study

RH1 - There is a significant change in the learning progression in terms of the achievement in ESL's BICS, CALP and overall scores of the learners in the successive trials of Class XI learners when Analytics-based Personalised Feedback is provided on the basis of performance.

3 METHODOLOGY

A quasi-experimental group design over multiple time points was used for the study. Therefore, the dependent variable of achievement in ESL was measured through the Achievement Test conducted four times in a single group at the interval of a month to collect data following time-series. A representative sample comprised of 39 students of Class XI, 11 girls and 28 boys of Kendriya Vidyalaya No. 1 who were selected randomly from the Bhopal city of M.P., India. The sample was taught by enrolling them in the online-course My Learning Class (MLC), using Moodle (Version 3.6.1) on the website learningenglish11.moodlecloud.com. Figure 1 presents the dashboard of the online course where learners enrolled to learn the intended content and take up the assessments in Google forms. Teacher used the platform to teach, assess the students and conduct different activities for the purpose.

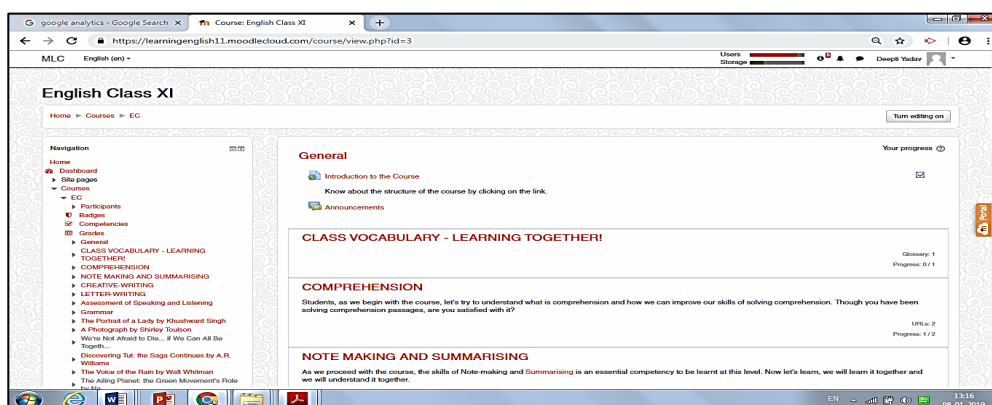


Figure 1: Dashboard of the Course at Moodle Platform for Class XI

Enrolled students were taught by the teacher using the Moodle LMS platform. Google forms were used to assess and provide AbPF during the treatment according to online performance. The feedback was provided to an individual learner on the basis of identifying their errors committed during the online performance to improve the next assessment attempt.

3.1 Tools and Statistical Techniques of the Study

An Achievement Test developed by the investigator was used to measure the learning outcomes in ESL, BICS and CALP. The Achievement Test items for BICS and CALP were finalised with the discriminating power 0.61 difficulty index of 0.70 calculated using Ebel's¹ (1972) approach. The data collected were analysed using Repeated Measures ANOVA.

4 RESULTS AND DISCUSSION

In the Figure 2, the Graph 1 (G 1) presents estimated marginal means BICS scores over the four tests out of 20 marks each; the Graph 2 (G 2) presents estimated marginal means CALP score over the four tests out of 30 marks each; and the Graph 3 (G 3) presents estimated marginal mean Overall score over the four tests out of 50 marks each. The first test scores have been used as the base for the comparison and map the progression.

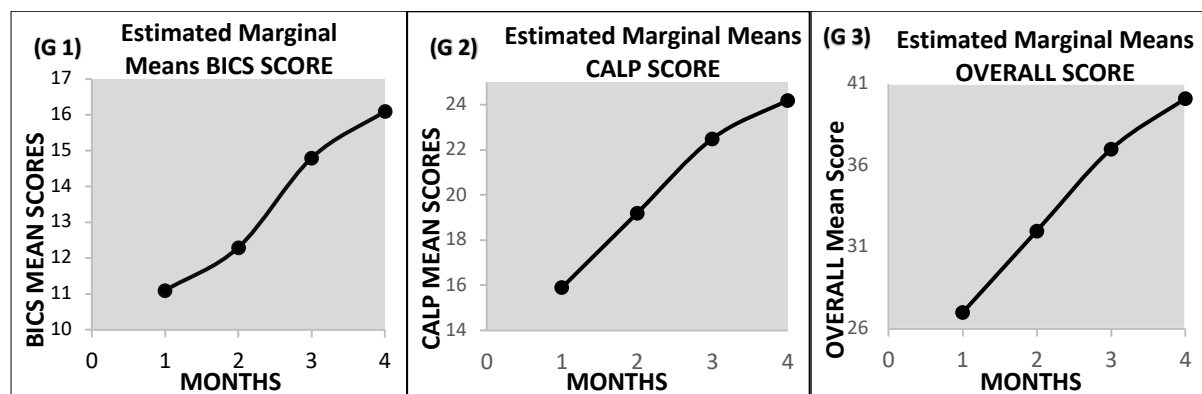


Figure 2: Estimated Marginal Means of BICS, CALP and Overall ESL Progression of Learners in the Successive Trials with Analytics-based Personalised Feedback

The first results of the ANOVA for BICS indicate that a significant time effect, Wilks' Lambda is 0.072 and F value of 154.40 is significant at 0.01 level with df equal to 3/37. The second results of the ANOVA for CALP indicate that a significant time effect, Wilks' Lambda is 0.065 and F value of 173.29 is significant at 0.01 level with df equal to 3/37. The third results of the ANOVA for Overall indicate that a significant time effect, Wilks' Lambda is 0.053 and F value of 213.651 is significant at 0.01 level with df equal to 3/37. Thus, it can be said that the alternative hypothesis namely, 'there is a significant change in the learning progression in terms of the achievement in ESL's BICS, CALP and Overall scores of the learners in the successive trials of Class XI learners when Analytics-based personalised feedback is provided on the basis of performance', is not rejected. The results show that the F-Value is significant for the BICS, CALP and Overall scores. So, it can be said that the scores in the different trials were not normally distributed. Follow-up comparisons indicate that each pairwise difference was significant, $p < 0.01$. There was a significant increase in scores over the time, suggesting that

¹ Difficulty index and discriminating power were calculated using Ebel's (1972) method. Based on the marks obtained, scores were rearranged descending order of magnitude, which was from highest to lowest. Then, the first 27 per cent and the last 27 per cent scores were used for item analysis, difficulty index and discriminating power using the formulas.

participation in the intervention for ESL, increased the learning outcomes level of BICS (Figure 2, Graph 1), CALP (Figure 2, Graph 2) and Overall (Figure 2, Graph 3). It can be inferred that the feedback provided to the learners brought a significantly positive effect on their learning outcomes of ESL.

5 FURTHER DELIBERATIONS, OBSERVATIONS AND FINDINGS

As the problem-areas were identified in the online performance assessment with Google forms, it was easier to guide with AbPF to the learners in their respective problem areas as shown in the Figure 3. There were twelve problem-areas identified in the

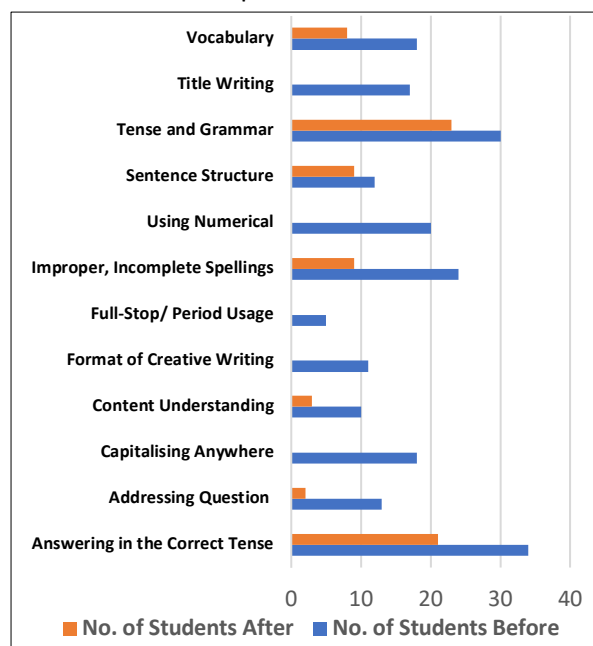


Figure – 3: Identified Problem-Areas for CALP and No. of Students Before and After the Treatment

performance, according to the online assessment of the performance for CALP competencies. These twelve problem-area components in the Figure 3 were the focus for feedback. For instance, a few students were writing the titles incorrectly and using numerical while writing English language. In another instance, while writing the closure of formal letter, 'Yours' was spelt incorrectly by many learners along with the incorrect format in the creative-writing. After providing them the AbPF, a constant deliberation to improve the identified problem-areas were worked upon by the teacher and learner. Likewise, all the identified problem areas presented in Figure 3 were dealt using AbPF.

Thus, a certain model eventually developed while using AbPF for the study has been presented in the Figure 4 below.

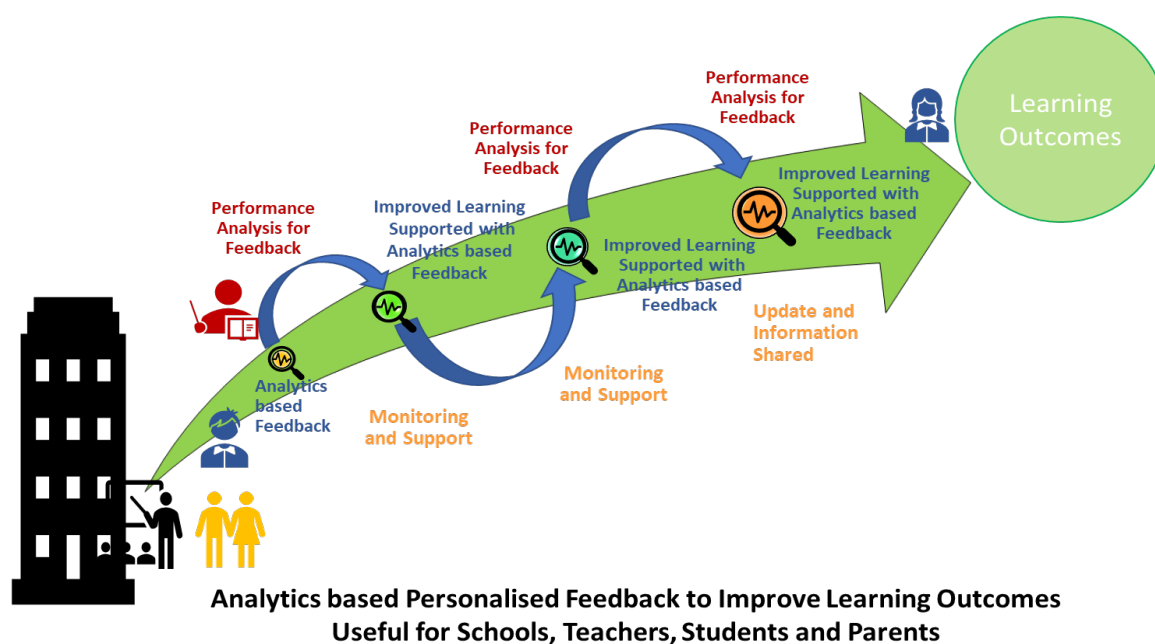


Figure 4: Analytics-based Personalised Feedback to Improve Learning Outcomes Useful for Schools, Teachers, Students and Parents

The above model explains the constant role of AbPF to reach the desired learning outcomes for a learner as well as teachers. According to the observations during the study, teacher's role becomes indispensable for the entire process, especially for language teaching. School and parents are also significant stakeholders supporting the learners by monitoring their online learning. Each feedback leads to some gain in understanding for the next step of learning to reach the goals. The study supports the outcome of Pardo et al. (2019) case study as the quality of the feedback matters for the academic achievement. Though Admiraal and Bulterman-Bos' (2017) case study could not produce noteworthy results but it significantly guided for the present study for an understanding among the stakeholders; and Peng (2017) had suggested to explore personalised teaching learning to teach foreign languages which the present study embarked on.

It was observed that when the learning progression for the two components of ESL, BICS and CALP increased significantly, there was an overall improvement in the ESL progression of the learners in successive trials with the timely AbPF. It signifies that AbPF has a significant effect on the learning outcomes of the Class XI learners in ESL for its two components BICS and CALP. It is implicated that the personalised feedback may be based on the identification of errors in their previous performance.

Learning Analytics may have the potential to change the face of education even individually, if employed with much needed planning and carefulness. Personalised feedbacks supported by Analytics can be significantly advantageous if the learner is perceiving it to improve the learning outcomes as observed in the present study, though much discussion and dialogues are needed on that perspective for a long-term implication.

REFERENCES

- Admiraal, W. & Bulterman-Bos, J.V.J. (2017). *Learning Analytics in Secondary Education: Assessment for Learning in 7th Grade Language Teaching*. ECER.
- Cummins, J. (1981). *Empirical and Theoretical Underpinnings of Bilingual Education*. Journal of Education, 163(1), 16-29.
- Liu, D. Y. T., Atif, A., Froissard, J. C., & Richards, D. (2019, January). *An Enhanced Learning Analytics Plugin for Moodle: Student Engagement and Personalised Intervention*. In *ASCILITE 2015- Australasian Society for Computers in Learning and Tertiary Education, Conference Proceedings*.
- Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., & Mirriahi, N. (2019). *Using Learning Analytics to Scale the Provision of Personalised Feedback*. British Journal of Educational Technology, 50(1), 128-138.
- Peng, Y. (2017). *Exploring Learning Analytics in College English Teaching and Learning in Big Data Era*. Atlantis Press, Volume 99, (501-512)
- Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). *The Current Landscape of Learning Analytics in Higher Education*. Computers in Human Behavior, 89, 98-110.

‘It’s a good tool, just hard to use’: Navigating staff perceptions during the institution-wide implementation of OnTask

Sheridan Gentili

UniSA, Teaching Innovation Unit
Sheridan.Gentili@unisa.edu.au

Amanda Richardson

UniSA, Teaching Innovation Unit
Amanda.Richardson@unisa.edu.au

ABSTRACT: There is universal agreement that feedback has a powerful influence on the academic success of students. Within the higher education sector, the delivery of high-quality feedback is increasingly impacted on by growing student class sizes and diminishing academic resources. User developed tools, such as OnTask, aim to address these issues. These types of platforms allow for large-scale creation and delivery of personalised data driven feedback based on students’ learning within their university learning management system (LMS). Yet, the adoption and actual use of such tools is often limited by staff perceptions of usefulness, and ease of use, as described by the Technology Acceptance Model. OnTask was piloted in 2017, made available to all academic staff across the university in 2018, with full LMS integration in 2019. While institution-wide user uptake has been low, the perceptions of those staff currently utilising the tool have been positive. This paper summarises the technology adoption process employed, describes current staff perceptions regarding the tool and discusses considerations around enabling broader adoption within the university.

Keywords: OnTask, Feedback, Learning Analytics, Technology Acceptance Model

1 INTRODUCTION

The contemporary higher education (HE) sector is faced with a number of challenges associated with improving student success and completion of higher education degrees. One of the key challenges faced by academic staff is the provision of just-in-time, meaningful and scalable feedback. Unfortunately, growing class sizes and diminished academic resources make it increasingly difficult to deliver personalised feedback to students. Various tools and models have been developed to minimise academics’ workload associated with the delivery of feedback. These include, but are not limited to, video feedback (Carruthers et al., 2015), and tools such as MetaTutor (Azevedo et al., 2012) and OnTask (Pardo et al., 2018). By utilising learning analytics (LA) and drawing on learner engagement data of specific learning activities within courses, these tools or models aim to improve student learning and reflection and are significant innovations for education.

In 2016 the Office of Learning and Teaching (OLT) funded a multi-institutional project aimed at addressing two key challenges faced by HE sector: 1) how to improve the quality of the student learning experience in the face of growing enrolments whilst 2) simultaneously increasing the maturity of LA deployment within the HE sector in Australia. The deliverables of the project included: 1) detailing the technical requirements for HE institutions; 2) guidelines to support academics in designing personalised feedback; and 3) the development of an open source platform for deployment within existing Learning Management Systems (LMS). The open source platform developed is called [OnTask](#). Within OnTask, academic staff (or instructors) are able to access course/subject learning and teaching data captured from with the LMS in real-time to generate

personalised feedback messages based on the learner data in a way akin to the ‘mail merge’ function in word processing tools (Figure 1).

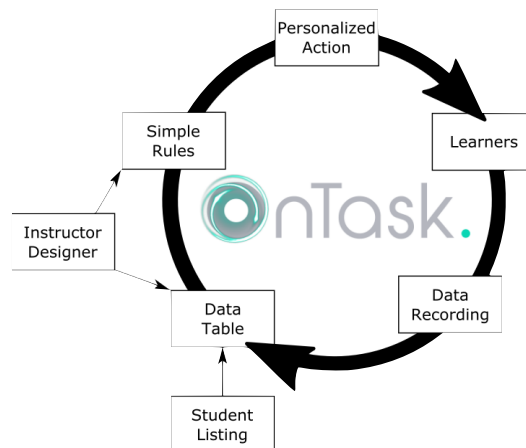


Figure 1. Summary of OnTask workflow used in generating personalised feedback.

The implementation of LA tools are intrinsically linked to the deployment of technology within the university setting. The operationalisation of such tools can be hampered at an institutional level by factors including data infrastructure, security concerns, technical support, and data governance issues (Colvin et al., 2017; Graf et al., 2012). Whilst an array of researcher led technologies like OnTask exist, they often fail to thrive in universities due to the complex nature of the HE sector (Colvin et al., 2017) and competing third party vendor software. Universities invest heavily in learning and teaching technologies, however academic staff often do not make full use of such tools for various reasons including ease of use, available time, and perceived relevance (Birch & Burnett, 2009; Shannon & Doube, 2004). The Technology Acceptance Model (Davis et al., 1989) based on the Theory of Reasoned Action (Sarver, 1983) summarises the complex interplay between an individual’s intent to perform a behaviour, their attitude towards the behaviour and social norms. The aim of this pilot study was to explore academic staff uptake of the OnTask tool and the challenges/successes experienced through a TAM lens following university-wide deployment across 2018 and 2019 to help inform the future direction, development and implementation of the tool across the university.

1.1 OnTask and the university-wide adoption strategy

OnTask was piloted in 2017 in selected undergraduate courses (Lim, Barker, Fudge, & Kelly, 2018; Lim et al., 2019). During that period the Teaching Innovation Unit (TIU), a central learning and teaching unit, worked to operationalise OnTask by providing full LMS integration. The TIU worked closely with the University’s IT unit to develop a shared understanding of how student learning and teaching data is captured in the LMS and stored. Full visibility of all LMS and student enrolment data was critical to ensure OnTask could be used as intended, as well as these data being accurately pushed to the OnTask tool. Seamless data integration required collaboration between several university units which posed several challenges.

Initially a subset of LMS data was available within OnTask, and regular quality control checks were performed to ensure the accuracy of the data. Incomplete or inaccurate data could lead to incorrect

feedback which may have a negative impact on students and effect staff confidence. Currently, data is pushed to OnTask from four separate data sources summarising LMS activity (transformed and live data) and lecture capture video data. These data are matched to current student enrolment data allowing academic staff to view a comprehensive profile of student online learning within their course. The data sources are pushed to OnTask every 24h and are constantly reviewed to ensure accuracy. Academic staff then access the data tables within OnTask to generate personalised feedback messages based on the learner data (Figure 1; Lim et al., 2018).

In 2018, academic staff were invited to participate in a soft-launch of OnTask, while in 2019 the tool was launched across the entire University with full application programming interface (API) integration with the LMS. Prior to the commencement of the major teaching period in March 2019, the TIU ran two staff development and training sessions. Fifty-eight staff from both the TIU (8 staff) and across the University (50 staff) attended, in addition to a virtual session attended by seven staff at regional campuses. These sessions were complemented by online help resources, including training videos developed by project partner institutions. To foster a community of practice, share ideas, and future enhancement of the tool, email distribution lists of all staff who attended the training sessions and a social networking group (using the enterprise software Yammer) were created. Anecdotally the tool appeared to be positively received, however uptake was low in the first half of 2019 (9% of all university courses, when split by mode of delivery 31% uptake in online courses and 6% uptake in face-to-face courses) with use ranging from once a week to a few times per teaching period. Unpacking the challenges faced by staff using OnTask is therefore important to both increase future uptake and enhance the tool.

2 EXPLORING STAFF EXPERIENCES THROUGH THE TECHNOLOGY ACCEPTANCE MODEL

The technology acceptance model (TAM) is used to understand how users accept and use new technology, providing valuable insights in how to improve institutional adoption and drive enhancements of the tool. Two fundamental determinants of whether an individual will accept and use technology is the perceived ease of use (PEOU) and perceived usefulness (PU) of the technology, which in this case relates to the impact on learning and teaching (Davis et al., 1989). PEOU is defined as the “the degree to which a person believes that using a particular technology would be free from effort” while PU is defined as “the degree to which a person believes that using a particular system would enhance his or her job performance” (Davis et al., 1989). An individual’s PEOU will impact directly on the PU of a tool, with both impacting on an individual’s attitudes towards using the technology (ATT), their behavioural intention to use (BI), and actual use (AU). ATT is defined as an individual’s feelings, either positive or negative, towards using technology, which is directly influenced by both PEOU and PU – if an individual perceives the technology as useful and easy to use, they tend to form positive emotions towards the technology itself. BI captures an individual’s conscious plans to make use of technology in the future, impacted on directly by the PU and ATT. These then shape a user’s actual use (AU).

To better understand the specific challenges experience by academic staff relating to their use of OnTask, we asked staff to reflect on the elements of the TAM. The brief questionnaire was composed of 9 questions across the 5 constructs (PEOU, PU, ATT, BI and AU; Table 1). Questions were adapted from prior studies (Shroff et al., 2011; Weng et al., 2018) and re-worded to meet the

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

specific context of this study. The nature of the responses did not allow for measures of internal consistency to be determined. Questions summarised in Table 1 were delivered via an anonymous survey using SurveyMonkey.

Table 1: Measurement items in the questionnaire.

TAM Construct	Question	Response Options
Perceived Ease of Use	How easy was OnTask to use?	Extremely easy / Moderately easy / Not at all easy
	How user-friendly is the OnTask software interface?	Extremely user-friendly / Very user-friendly / Moderately user-friendly / Not very user-friendly / Not user-friendly at all
	Were the support resources helpful?	Very helpful / Sometimes / Not helpful at all
	In your view, what areas need to be improved?	Free text
Perceived Usefulness	Do you believe the learning and teaching experience of your students was enhanced by the use of OnTask?	Definitely / Neither yes or no / It made no impact
Attitudes Towards Using	Would you recommend the use of OnTask to your colleagues, and if so why	Yes / No
Behavioural Intent to Use	Will you use OnTask again?	Yes / No
Actual Use	Did you use OnTask in your course?	Yes / No
	What type of changes did you make to your course to enable you to use OnTask?	Significant number of changes / Few changes / No changes

2.1 Staff perceptions of OnTask - TAM results

Of the 50 university-wide staff who attended the face-to-face TIU training sessions, 22% of staff responded to the survey with representation from each of the university's academic units (Business, 36.4%; Arts & Education, 9.1%; Health, 27.3%; IT & Engineering, 9.1%; Other 18.2%).

2.2 Perceived Ease of Use (PEOU) and Perceived Usefulness (PU)

Of the staff surveyed, 100% were actively using the tool in 2019 and 82% found the tool moderately easy to use, whilst 18% found the tool difficult to use. Interestingly, of those staff who found the tool moderately easy to use, 67% indicated that the user interface wasn't very user-friendly, while 33% believed it to be moderately to very user-friendly. The staff that found OnTask difficult to use were evenly split regarding their perceptions of the user-friendliness of the interface. Independent of staff views with regards to user-friendliness of the interface, 90% of staff believed that the OnTask email had a positive impact on undergraduate student learning, with 10% uncertain of the impact. Staff opinions of the resources developed by the TIU varied - 18% found them to be very helpful, 72%

found them to be helpful at times, whilst 9% found them to be less than helpful. Interestingly the same 9% reported the OnTask user interface as not being user-friendly.

2.3 Attitudes Towards Using (ATT), Behavioural Intention to Use (BI) and Actual Use (AU)

The majority of staff surveyed (70%) made no changes to their undergraduate course prior to the deployment of OnTask. Of those who made changes (30%) all cited the changes as being minor to improve the type of data being pushed to OnTask.

Despite the PEOU and PU outcomes, there was no relationship between the helpfulness of resources and the likelihood of staff using OnTask again, with 91% of staff surveyed citing that they would use the tool when they next taught. Consistent with this, 90% of staff would recommend the use of OnTask to colleagues. Concerns, however, were raised by 64% of staff regarding the reliability of data. Whilst the University has quality control processes in place to ensure the data captured and transformed is accurate, the transfer of data to OnTask is impacted on by a range of institutional IT factors beyond the control of the TIU. The main improvement requested by staff related to the user interface with staff believing the interface to be “clunky and not user-friendly”.

2.4 Summary of Find findings and Future Steps

The Perceived Ease of Use (PEOU) of the OnTask tool was reasonably high despite negative staff views regarding the user-friendly nature of the tools interface, with one member of staff commenting “It is a good tool just difficult to use”. Consistent with this statement, staff Perceived Usefulness (PU) of the tool was high with staff reporting the tool as being “a great way to connect with students on a more individual level”. Furthermore, staff indicated that students “commented and appreciated individualised correspondence”, positively impacting on Attitudes Towards Using (ATT), Behavioural Intention to Use (BI) and Actual Use (AU) of OnTask.

Successful adoption of learning and teaching technologies as add-ons to the existing LMS requires strong instructional support and prioritisation as part of a learning and teaching strategic plan (King & Boyatt, 2015). Innovators and early adopters are likely to engage with new technologies early, seeing the relevance and overlooking early failures in support and software performance, however late majority and laggards tend not to engage with such tools unless incentivised (Porter & Graham, 2016; Rogers, 2003). Although the implementation of OnTask occurred at a university level coordinated by a central unit, OnTask does not currently form part of the University’s learning and teaching strategy. Whilst the use of OnTask across the institution is growing, this use is closely linked to research projects focussed on the quality and the impact of feedback on student learning (Lim et al., 2019, 2018). Positive staff experiences, as reported in this paper, increase the capacity for central units to engage with senior management to incorporate such tools into strategy, increasing the adoption of the tool beyond the research focused academic group. The articulation of staff PU of the tool with the broader priorities of the university allows for more seamless integration into strategy, subsequently incentivising the use of the tool. The critical role of leadership in the sustained implementation of LA tools is well recognised, specifically a deep scholarly understand of LA processes by leaders as factors in uptake and integration (Colvin et al., 2017). A strategic top-

down bottom-up approach to the integration of learning tools, like OnTask, will increase the likelihood of wide-spread adoption.

The reliability of the data being pushed to OnTask was the commonly reported concern for staff. Data accuracy is essential given that OnTask student feedback is generated using the data available within the tool (Pardo et al., 2018). Delays or failures in data transfer may cause inaccuracies in the feedback messages generated, negatively impacting on student perceptions of the feedback, and staff perceptions of the technology. Student responses to feedback can often be defensive, particularly when feedback provides a critique of submitted work or contains low grades (Pitt & Norton, 2017; Robinson et al., 2013), inaccurate messaging has the potential to exacerbate these feelings. To minimise the likelihood of a negative student response, it is critical that student feedback is generated using accurate data. From a staff perspective, early work in the field of LA suggests that academics can see the benefits of LA in improving learning and teaching, however a sufficient level of trust associated with data used is required (Drachsler & Greller, 2012). In the context of this paper, delays or data errors can lead to a mis-trust in the tool as a resource to support learning and teaching - an important consideration when aiming for institution wide adoption.

Although staff noted concerns around data reliability, these appeared to have little impact on their intent to use (BI) and actual use (AU) of OnTask. Important to note though is that the staff who responded to the survey and are already using OnTask are likely to be the innovators and early adopters (Rogers, 2003) within the whole university staff cohort. Implementation of OnTask at an institutional level would require buy-in across all types of innovation adopters, i.e. early majority, late majority and laggards accounting for approximately 84% of academic staff (Rogers, 2003). In the absence of an institutional strategy, negative views associated with the reliability of data may contribute to a reduction in wide spread adoption of OnTask across the remainder of academic staff.

3 CONCLUSIONS

Key drivers for the development and use of learning technologies in higher education include the improvement of students' learning experiences whilst simultaneously meeting broader institutional needs. User developed tools such as OnTask should be used to re-think learning and teaching in HE in ways to improve the delivery of student feedback, known to significantly impact on student success (Carless & Boud, 2018). Integration of such systems within existing university structures creates challenges around how disconnected units within an institution must work together to seamlessly embed system wide processes to achieve a common goal. Ensuring student data captured by LMS are accurate and available in real-time will help build staff confidence, but more importantly will allow for effective and appropriate personalised feedback to be delivered to students. Furthermore, senior management support is required to ensure articulation of such tools into institutional strategies, thereby motivating academic staff to engage with the technology.

REFERENCES

Azevedo, R., Bouchet, F., Harley, J., Feyzi Behnagh, R., Trevors, G., Duffy, M., Taub, M., Pacampara, N., Agnew, L., Griscom, S., Mudrick, N., Stead, V., & Yang, W. (2012, June 1). MetaTutor: An

- Intelligent Multi-Agent Tutoring System Designed to Detect, Track, Model, and Foster Self-Regulated Learning. <https://doi.org/10.13140/RG.2.1.1334.6640>
- Birch, D. T., & Burnett, B. M. (2009). Bringing academics on board: Encouraging institution-wide diffusion of e-learning environments. <https://doi.org/10.14742/ajet.1184>
- Carruthers, C., McCarron, B., Bolan, P., Devine, A., McMahon-Beattie, U., & Burns, A. (2015). 'I like the sound of that' – an evaluation of providing audio feedback via the virtual learning environment for summative assessment. *Assessment & Evaluation in Higher Education*, 40(3), 352–370. <https://doi.org/10.1080/02602938.2014.917145>
- Colvin, C., Dawson, S., Wade, A., & Gašević, D. (2017). Addressing the challenges of institutional adoption. *Handbook of Learning Analytics*, 281–289.
- Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, 35(8), 982–1003. <https://doi.org/10.1287/mnsc.35.8.982>
- Drachsler, H., & Greller, W. (2012). Confidence in Learning Analytics. *Educational Technology & Society*.
- Eckhouse, B., & Carroll, R. (2013). Voice Assessment of Student Work: Recent Studies and Emerging Technologies. *Business Communication Quarterly*, 76(4), 458–473. <https://doi.org/10.1177/1080569913506488>
- Graf, S., Ives, C., Lockyer, L., Hobson, P., & Clow, D. (2012). Building a data governance model for learning analytics. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 21–22. <https://doi.org/10.1145/2330601.2330614>
- King, E., & Boyatt, R. (2015). Exploring factors that influence adoption of e-learning within higher education. *British Journal of Educational Technology*, 46(6), 1272–1280. <https://doi.org/10.1111/bjet.12195>
- Lim, L.-A., Barker, S., Fudge, A., & Kelly, S. (2018). Keeping everyone OnTask: Gauging the impact of personalised feedback through academic case studies. *Open Oceans: Learning without Borders.*, 184–193.
- Lim, L.-A., Gentili, S., Pardo, A., Kovanović, V., Whitelock-Wainwright, A., Gašević, D., & Dawson, S. (2019). What changes, and for whom? A study of the impact of learning analytics-based process feedback in a large course. *Learning and Instruction*. <https://doi.org/10.1016/j.learninstruc.2019.04.003>
- Pardo, A., Bartimote, K., Shum, S. B., Dawson, S., Gao, J., Gašević, D., Leichtweis, S., Liu, D., Martínez-Maldonado, R., Mirriahi, N., Moskal, A. C. M., Schulte, J., Siemens, G., & Vigentini, L. (2018). OnTask: Delivering Data-Informed, Personalized Learning Support Actions. *Journal of Learning Analytics*, 5(3), 235–249–235–249. <https://doi.org/10.18608/jla.2018.53.15>
- Pitt, E., & Norton, L. (2017). 'Now that's the feedback I want!' Students' reactions to feedback on graded work and what they do with it. *Assessment & Evaluation in Higher Education*, 42(4), 499–516. <https://doi.org/10.1080/02602938.2016.1142500>
- Porter, W. W., & Graham, C. R. (2016). Institutional drivers and barriers to faculty adoption of blended learning in higher education. *British Journal of Educational Technology*, 47(4), 748–762. <https://doi.org/10.1111/bjet.12269>
- Robinson, S., Pope, D., & Holyoak, L. (2013). Can we meet their expectations? Experiences and perceptions of feedback in first year undergraduate students. *Assessment & Evaluation in Higher Education*, 38(3), 260–272. <https://doi.org/10.1080/02602938.2011.629291>
- Rogers, E. M. (2003). *Diffusion of Innovations* (5th ed.). Simon and Schuster.

- Sarver, V. T. (1983). Ajzen and Fishbein's "Theory of Reasoned Action": A Critical Assessment. *Journal for the Theory of Social Behaviour*, 13(2), 155–164. <https://doi.org/10.1111/j.1468-5914.1983.tb00469.x>
- Shannon, S., & Doube, L. (2004). Valuing and using web supported teaching: A staff development role in closing the gaps. *Australasian Journal of Educational Technology*, 20. <https://doi.org/10.14742/ajet.1370>
- Shroff, R. H., Deneen, C. C., & Ng, E. M. W. (2011). Analysis of the technology acceptance model in examining students' behavioural intention to use an e-portfolio system. *Australasian Journal of Educational Technology*, 27(4). <https://doi.org/10.14742/ajet.940>
- Watty, K., McKay, J., & Ngo, L. (2016). Innovators or inhibitors? Accounting faculty resistance to new educational technologies in higher education. *Journal of Accounting Education*, 36, 1–15. <https://doi.org/10.1016/j.jaccedu.2016.03.003>
- Weng, F., Yang, R.-J., Ho, H.-J., & Su, H.-M. (2018). A TAM-Based Study of the Attitude towards Use Intention of Multimedia among School Teachers. *Applied System Innovation*, 1(3), 36. <https://doi.org/10.3390/asi1030036>

Social Network-based Awareness Tools in Collaborative Learning

Lei Zhang, Xiangdong Chen

Department of Education Information Technology, East China Normal University
azhangyilei@163.com, xdchen@deit.ecnu.edu.cn

ABSTRACT: Although various types of CSCL awareness tools are provided during collaboration, few research is done to integrate social network feedback into collaborative learning. The purpose of the study, therefore, is to examine the incorporation of social network analysis-based awareness tools in collaborative learning. In this study, four types of social network awareness tools were developed and utilized: 1) whole network representation, 2) ego network representation, 3) position analysis, 4) participation representation. A small-scale case study is conducted to illustrate the way social network analysis-based awareness tools that can be used in collaborative learning. Impacts of those social network awareness tools were explained and students' perceptions of those awareness tools were discussed.

Keywords: social network analysis, collaborative learning, group awareness

1 INTRODUCTION

Group awareness plays an important role in the quality of group collaboration (Järvenoja, Järvelä, & Malmberg, 2017). With the development of CSCL, an increasing number of researchers conducted studies related to collaborative learning awareness tools (Buder & Bodemer, 2008; Järvenoja, Järvelä, & Malmberg, 2017). Many researchers utilized different visualization methods, including radar chart (Phielix, Prins, & Kirschner 2010), scatter plot (Buder & Bodemer, 2008), and tables (Miller & Hadwin, 2015) as collaborative learning awareness tools. However, these studies had limitations. Firstly, current collaborative learning awareness tools neglected social interactions in collaborative learning (Cho, Gay, Davidson & Ingraffea, 2007). Collaborative learning involves a lot of social interactions among members (Heo, Lim, & Kim, 2010), but there is a lack of study that provides social interaction feedback for learners (De Laat, Lally, Lipponen, & Simons, 2007). How group members communicate with others? Who is the leader in the group? Who is neglected in the group? These questions are unanswered. Secondly, the efficiency to utilize available data is low. Although a great amount of process data, including online blogs, log files, videos, is collected, the social interactions (e.g. face expression, mood) contained in that data is little utilized.

Thus, there is a need to integrate social network analysis into collaborative learning awareness tools. A series of social network analysis-based awareness tools (SNA-based awareness tools) in this study were created. These tools were used in a small-scale study. Effects of these tools were discussed. This study helps to guide better practices for researchers who are interested in using social network-based awareness tools.

2 METHODS

This paper is based on a small-scale study. It examines the application and impacts of SNA-based awareness tools in a collaborative learning context. A total of 26 undergraduates (5 groups, 4-6 people in a group) in one university located in the eastern part of China participated in this study. They enrolled in the same course, Information Technology Teaching Method. In this class, they need to collaborate with each other to write a lesson plan. Groups' collaboration processes were videotaped and online discussions from the online platform we developed were collected in this course. A social network questionnaire was also utilized to collect information about students' social interactions. All data was analyzed with social network analysis methods and visualized using the social network-based awareness tools on the online learning platform. In addition, interviews about students' perceptions on those SNA-based awareness tools were conducted. A five-Likert survey was also used to investigate students' perceptions about the collaboration quality after class. This survey was designed based on previous studies (Lee, 2014). There are two dimensions in the survey: social interactions (4 items) and participation (8 items).

3 RESULTS

Before collaboration, some students may be familiar with each other, and a whole network social gram is necessary for them to know the social relationships in this class and seek for group members. Thus, students' basic information (name, gender, previous collaborating experience with others) were collected and a whole network social gram (Figure 1a) was provided for students to help them form a group. To protect privacy, we used number (1,2,3.....) to represent participants in this paper. For instance, before collaboration, member 1, 2, 3 were friends and they had collaboration experience before this class; member 4, 5, 6 were not familiar with others.

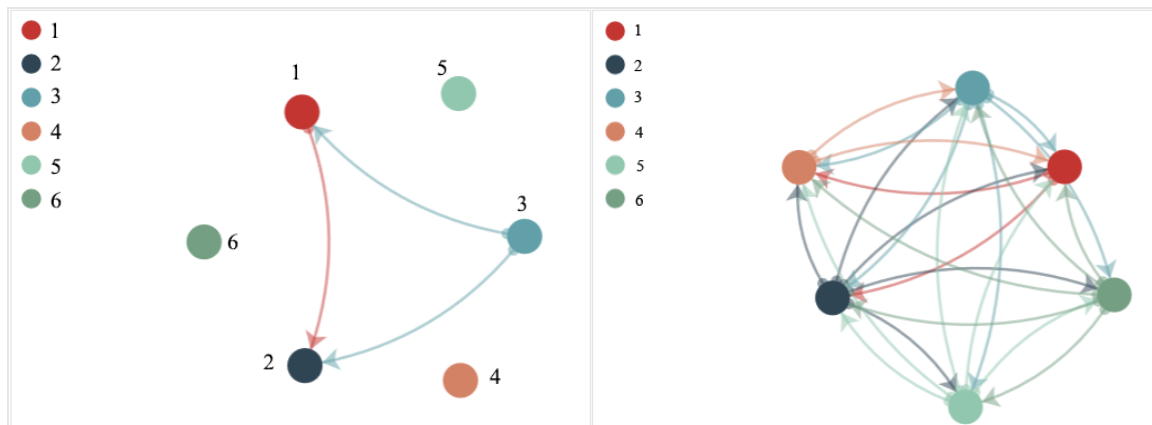


Figure 1a: whole network social gram

Figure 1b: whole network social gram

While collaborating, group members regulated with each other to write a lesson plan together and three types of social network awareness tools were provided: (1) ego network representation, (2) position representation, (3) participation representation. Different groups were provided with personalized SNA-based graphics according to their different contributions and social structures.

- (1) Ego network representation. This visualization tool aims to provide information about members' own interactions with others. It is analyzed using in and out degree. Take group two for instance,

figure 2 showed the ego network of a member (member 1) in group two. There were six members in group 2 (1,2,3,4,5,6). Each circle represents a member in this group. During collaborative learning, member 1 interacts mutually with member 3 and member 4 (e.g. member 1 communicate with member 3 and member 4); member 5 has a one-way communication with member 1 (e.g. member 5 reminds member 1 to submit task). Thus, the indegree of member 2 is three, and the out degree of member 1 is two.

- (2) Position representation. This visualization aims to show group members' position. This is analyzed by position analysis. Take group two for example, figure 3 shows group members' position or role in a group. Compared with other members, member 2 were in a leader position when monitoring group's progress.
- (3) Participation representation. This visualization aims to represents each member's participation. It is analyzed using 2-mode network. Figure 4 represents a participation representation (rectangles represented events, circles represented group members). It visualizes the degree to which different members participated in different events (e.g. task understanding, goal setting, etc.). If the rectangle is larger, that means the group spend more time in this event; If the circle is larger, that means this member contribute more in a group.

Similar to figure 1a, after collaboration, a whole network social gram was also provided for groups, indicating their social interactions with others after collaboration. The data from online discussions that each group posted to the online platform was collected and the post-and-reply relations were analyzed using social network analysis. Then a whole network social gram was represented for students (Figure 1b).

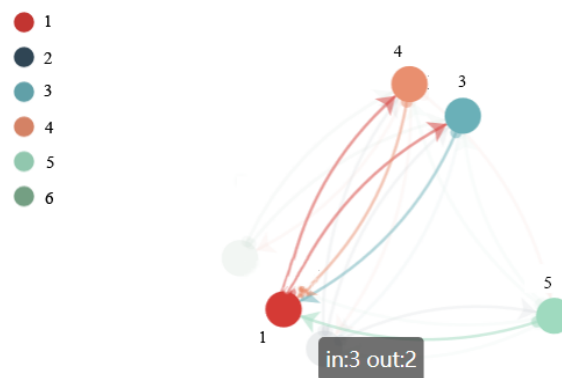


Figure 2: ego network representation (group 2)

Who will monitor task progress?

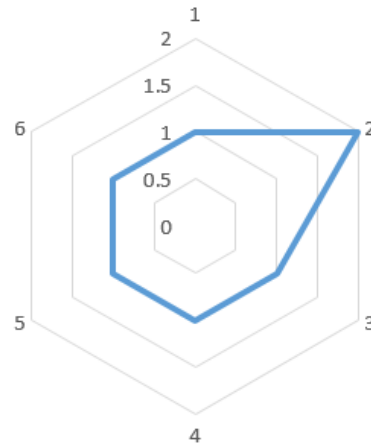


Figure 3: position representation (group 2)



Figure 4: participation representation (group 2)

4 FINDINGS

4.1 Collaboration quality

To exam the impact of using those SNA-based awareness tools, we analyzed groups' communication network. Results revealed that both density (Table 1) and degree (Table 2) of each group' communication network increase, indicating that more members in the same group interacted with others, and these tools were effective in promoting the quality of collaboration.

Table 1: Density of each group

	Group 1		Group 2		Group 3		Group 4		Group 5	
	1	2	1	2	1	2	1	2	1	2
Density	0.33	0.6	0.50	0.7	0.46	0.6	0.47	0.53	0.50	0.67

Note. 1=before collaboration, 2=after collaboration

Table 2: Degree of each group

	Group 1		Group 2		Group 3		Group 4		Group 5	
	1	2	1	2	1	2	1	2	1	2
Degree	1.67	3.0	2.0	2.8	1.2	2.4	1.67	2.67	1.5	2.0

Note. 1=before collaboration, 2=after collaboration

Furthermore, by analyzing the five-Likert survey data we collected after collaboration, we found that the overall collaborative quality remained high, 4.39 in average. In addition, the average scores in social interaction dimension and participation dimension were 4.43 and 4.35 respectively.

4.2 Students' perceptions of using SNA-based tools

In addition, interviews with different group members also showed that the use of this SNA-based awareness tools was useful. On the one hand, these tools helped students to self-assess their own participation or behaviors. Many students expressed that they will reflect what they have contributed after seeing these figures. Below is one student's feelings after seeing the position representation.

"I would pay attention to the difference between me and other members, especially the one who contributed more. I will reflect what I did during collaboration and try to refine my behaviors in the future."

On the other hand, these tools also helped group members to regulate with each other. Out of 10 students we interviewed, all of them agreed to be more aware of group members' participation. They would keep tabs on peers who were neglected and pay more attention to them. Furthermore, 8 out of 10 students pointed out that they will reflect their own behaviors after seeing those social network graphics, then regulated their behaviors to complete group tasks.

"We will discuss together after seeing these pictures.....We will reflect together and try to refine our group's collaboration."

5 CONCLUSION

In this study, we conducted a case study to examine the way in which social network analysis-based awareness tools can be used in collaborative learning. Four types of social network-based awareness tools are developed, which highlight best practice and guide researchers who are interested in using this type of awareness tool. This study adds to the discussion of social network-based awareness tool in collaborative learning and contributes to the understanding of social network-based awareness tools practice.

REFERENCE

De Laat, M., Lally, V., Lipponen, L., & Simons, R. J. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

- Network Analysis. *International Journal of Computer-Supported Collaborative Learning*, 2(1), 87-103.
- Håkansson, H., Havila, V., & Pedersen, A. C. (1999). Learning in networks. *Industrial marketing management*, 28(5), 443-452.
- Hatala, J. P., & Fleming, P. R. (2007). Making transfer climate visible: Utilizing social network analysis to facilitate the transfer of training. *Human Resource Development Review*, 6(1), 33-63.
- Heo, H., Lim, K. Y., & Kim, Y. (2010). Exploratory study on the patterns of online interaction and knowledge co-construction in project-based learning. *Computers & Education*, 55(3), 1383-1392.
- Janssen, J., Erkens, G., & Kirschner, P. A. (2011). Group awareness tools: It's what you do with it that matters. *Computers in human behavior*, 27(3), 1046-1058.
- Järvelä, S., Kirschner, P. A., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., ... & Järvenoja, H. (2015). Enhancing socially shared regulation in collaborative learning groups: designing for CSCL regulation tools. *Educational Technology Research and Development*, 63(1), 125-142.
- Lee, A. (2014). *Socially shared regulation in computer-supported collaborative learning* (Doctoral dissertation, Rutgers University-Graduate School-New Brunswick).
- Nurmela, K., Lehtinen, E., & Palonen, T. (1999, December). Evaluating CSCL log files by social network analysis. In *Proceedings of the 1999 conference on Computer support for collaborative learning* (p. 54). International Society of the Learning Sciences.
- Phielix, C., Prins, F. J., & Kirschner, P. A. (2010). Awareness of group performance in a CSCL-environment: Effects of peer feedback and reflection. *Computers in Human Behavior*, 26(2), 151-161.
- Saqr, M., Fors, U., Tedre, M., & Nouri, J. (2018). How social network analysis can be used to monitor online collaborative learning and guide an informed intervention. *PloS one*, 13(3), e0194777.
- Van den Bossche, P., Segers, M., & Jansen, N. (2010). Transfer of training: the role of feedback in supportive social networks. *International Journal of Training and Development*, 14(2), 81-94.

Toward the Design of a Learning Analytics for Learning Design Dashboard in Location-based Learning

Gerti Pishtari, María Jesús Rodríguez-Triana

Tallinn University

gerti.pishtari@tlu.ee, mjrt@tlu.ee

ABSTRACT: This paper presents the design process of a set of analytical dashboards for Rada, a location-based authoring tool that allows the design of geo-localized learning scenarios in open-air environments. The dashboards can support practitioners in their learning design practices outside the classroom, as well as to provide real-time and personalized feedback to students. We propose a design process that combines a bottom-up approach by involving different stakeholders and design sessions, with a top-down approach of evidence from good practices found in the literature, done through a systematic review. Lessons learned from the process emphasize the added value of involving the community of stakeholders behind the tool during the entire design process.

Keywords: Location-based Authoring Tool, Dashboards, Learning Design, Multi-stakeholder Analytics, Learning Analytics

1 INTRODUCTION

Location-based authoring tools allow the design of geo-localized learning experiences outside the classroom. Practitioners can make use of such authoring tools to transform specific situated environments (e.g., a thematic park, or museum), into technology-enhanced learning environments, in line with their pedagogical goals. However, such technologies also entail added complexity, e.g., evaluating learning scenarios and providing feedback in these settings requires practitioners to make sense of learning activities that happen across digital and physical spaces. Moreover, the distributed nature of learning outside the classroom also impacts the ability that practitioners have to provide feedback to their students, during, as well as after the learning activity (Pishtari et. al, 2019a).

The field of learning analytics, more specifically the attempts to align learning design and learning analytics practices (Lockyer et. al, 2013), as well as recent trends toward multimodality (Ochoa, 2017), could help to address these issues. In the case of location-based learning, aligning learning design and learning analytics can help practitioners to design and keep track of learning activities that include both formal and non-formal learning elements (e.g., an activity that partially happens in a classroom, and partially outdoor). Furthermore, integrating the information provided from different data sources (multimodal learning analytics) could help to keep practitioners informed in real-time about students' activities and performance, hence enhance their ability to provide feedback in a location-based learning environment.

This paper presents the design process of a set of dashboards for Rada¹, a location-based authoring tool that allows the design of geo-localized learning activities. To develop the dashboards we take a multi-stakeholder approach, by considering several stakeholders that are interested in better understanding learning design practices, as presented in Pishtari et al, 2019b. However, in this paper we mainly focus on the design indicators and intend to help practitioners to reflect on the activities, as well as to provide feedback to students, both during the activity (in real-time) and after it.

2 CONTEXT OF THE RESEARCH

The process presented in this paper has been carried out in the context of Rada, a location-based authoring tool that allows practitioners to create and conduct learning activities with game elements outside the classroom, as tracks. Each track consists of a number of location points where specific tasks are assigned. Practitioners can create learning activities based on a list of templates with different tasks (such as single/multiple correct answers, match pairs, freeform answers, etc.), which can be freely assigned to a location point on the map. While playing the game, location points are activated when students reach close enough to the specific location and change color once the answer is submitted. Depending on how the practitioner has structured the track of tasks, students might have to follow a predefined order to respond to the location points, or randomly. Rada also provides immediate feedback to students for each submitted answer and awards achievements with points and badges. When the track is over, students can access the overall results. Currently, Rada displays average data on a dashboard about created tracks that include places where tracks have been created, time spent on a specific track, the number of correct and wrong answers, etc.

3 DESIGN PROCESS

To design the dashboards, we combined a bottom-up approach of involving stakeholders through a cycle of interviews and design iterations, with a top-down approach of conducting a literature review (see Table 1). In the first phase, we conducted a systematic literature review about learning design and learning analytics in mobile and ubiquitous learning (Pishtari et al, 2019a). The results from this review emphasize the benefits that aligning learning design and learning analytics in these settings has on practitioners' practices, especially supporting evidence-based decision making and providing contextualized and personalized feedback. Simultaneously, in order to better understand stakeholders needs, on how learning analytics can support the processes of learning design, monitoring, and providing feedback in location-based learning, we organised a set of contextual inquiries with five practitioners, as well as semi-structured interviews with two researchers and two managers (Pishtari et al, 2019b). Results obtained from the interviews were analysed qualitatively (see Table 1), and later organized according to the AL4LD framework (Hernández-Leo et. al, 2019). This framework structures the support that analytics can provide learning design into different layers, by taking a multi-stakeholder approach. We also grouped the results into the ones that are general in the context of analytics for learning design, and the ones that are specific to the context of location-based learning. Results connected to the context of location-based learning specifically emphasize the need for real-time monitoring and the possibility to communicate the personalized feedback to students during the learning activities (Pishtari et. al, 2019b).

¹<http://web.htk.tlu.ee/rada>

The second phase consisted in a design session, similar to a design sprint². Six experts in the field of technology enhanced learning (TEL), but no previous experience with Rada, participated in a design sprint. The goal of this session was to brainstorm and produce design ideas that could address the needs that were identified in the previous phase. The design session consisted of four phases: a) understanding the context, during which we organized a short presentation of the tool and the results from the previous phase; b) reviewing and remixing existing ideas as well as sketching possible solutions; c) evaluating the proposed solution, selecting the potential solutions and designing storyboards aligned with each solution; d) transform the ideas from the storyboard into prototypes. Three main metaphors resulted from the design session. For each of them we designed a separate and very different dashboard as a paper prototype³. We called the first metaphor **zoom in**, because its dashboard allows practitioners to zoom into specific details of the data (e.g., for a specific student, during a specific task and moment), or zoom out and maintain a general overview of the ongoing/finished activity. The second metaphor was called **timeline**, as it allows practitioners to follow what is happening in the activity through a newsfeed that organized the happenings as events. The third metaphor was connected to the other community stakeholders and we called it **reporting**, as it allows the user to select specific variables, or sources of information from the data available (usually based on a set of predefined questions that the user has). The dashboard will allow the user to organize the information as a report with results and conclusions.

The third and last phase of the design sprint was the evaluation. This is still an ongoing process, for which we have evaluated the dashboards with users that had already used Rada before. The particularity of this evaluation is that the information shown on the dashboards is based on real data from sessions that the users have designed, which makes the feedback provided more contextualized. Results from this phase will be analysed contextually and will result in the first usable prototypes. The next steps towards the evaluation will consist of pilots that will make use of these real prototypes, where students will also be included.

Table 1: Summary of the design process, participants and the techniques used.

Phase	Event	Participants	Data-analysis technique
Phase 1	Systematic literature review	-	- Qualitative and quantitative
	Contextual inquiry	- 5 Practitioners	- Interpretation sessions - Affinity diagramming - AL4LD framework
	Semi-structured interview	- 2 Researchers - 2 Managers	- Thematic analysis - AL4LD framework
Phase 2	Design sprint	- 6 TEL Experts	- Co-design session
Phase 3	Paper prototype evaluation	- 4 Practitioners - 2 Researchers - 1 Developer	- Content analysis

² <https://www.gv.com/sprint/>

³ Paper prototypes versions of each dashboard: <http://bit.ly/RadaDashboardsMockups>

4 CONCLUSION

This paper presents the design process of a set of analytical dashboards in Rada, that aim to support practitioners in their learning design practices outside the classroom, as well as to provide evidence-based feedback to students. During the process we take a multi-stakeholder approach, as we consider that informing community stakeholders such as researchers and managers of educational institutions is crucial in understanding the usage, impact and adoption of such tools.

To achieve these goals, we propose a combination of a top down approach based on literature, with a bottom-up approach of co-designing together with related stakeholders. The design process proposed in this paper is composed of a set of contextual inquiries and interviews with stakeholders, a design session that can guide the transformation of the input from the interviews into concrete ideas or prototypes, and an evaluation procedure with real data. The proposed process could be used to guide the conceptualization and deployment of learning analytics dashboards that take into consideration the needs of specific stakeholders during the design process.

In upcoming design iterations, we are planning to integrate the information gathered from the evaluations into Rada and organise a set of pilot evaluations in real-settings, which will involve students as well. Through this pilot we plan to gather further evidence about the support that analytics could provide to learning design practices, as well as to the process of monitoring and providing personalized feedback in location-based learning.

ACKNOWLEDGMENTS

This research has been partially funded by the European Union in the context of CEITER (Horizon 2020 Research and Innovation Programme, grant agreement no. 669074).

REFERENCES

- Hernández-Leo, D., Martínez-Maldonado, R., Pardo, A., Muñoz-Cristóbal, J. A., & Rodríguez-Triana, M. J. (2019). Analytics for learning design: A layered framework and tools. *British Journal of Educational Technology*, 50(1), 139-152.
- Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist*, 57(10), 1439-1459.
- Pishtari, G., Rodríguez-Triana, M. J., Sarmiento-Márquez, E. M., Terasmaa, J., Kori, K., Kangur, M., ... & Puusepp, L. (2019a, September). An Overview of Learning Design and Analytics in Mobile and Ubiquitous Learning. In *International Conference on Web-Based Learning* (pp. 312-319). Springer, Cham.
- Pishtari, G., Rodríguez-Triana, M. J., & Väljataga, T. (2019b, September). Multi-stakeholder Analytics for Learning Design: A Case Study of Location-based Tools. In *World Conference on Mobile and Contextual Learning* (pp. 94-101).
- Ochoa, X. (2017). Multimodal learning analytics. *The Handbook of Learning Analytics*, pp.129-141.

Learning Analytics Support for Dialogic Peer Feedback

Erkan Er

Universidad de Valladolid
erkan@gsic.uva.es

Yannis Dimitriadis

Universidad de Valladolid
yannis@tel.uva.es

Dragan Gašević

Monash University
dragan.gasevic@monash.edu

ABSTRACT: Dialogic peer feedback can be an effective approach to the design of scalable and impactful feedback practices in higher education. However, the literature lacks a clear framing of dialogic peer feedback from a theoretical perspective, which is necessary for a solid and systematic design of feedback practices. This paper introduces a theoretical framework which suggests a collaboration among peers during three phases for a successful feedback activity. Then, the Synergy platform is introduced. Synergy, designed to facilitate dialogic peer feedback, is grounded in the presented framework. Instructor facilitation is necessary to support students in various learning processes during dialogic peer feedback. Synergy includes various LA components to support instructor actions. The potential of LA support to offer actionable insights for instructors' timely intervention is briefly discussed.

Keywords: learning analytics, dialogic peer feedback, learning analytics dashboards

1 INTRODUCTION

Feedback often produces lesser learning gains than expected in higher education. Dialogue can increase the impact of feedback (Yang & Carless, 2013). However, there is a need for a systematic design when dialogue is integrated into peer feedback. Most of the existing work optimistically rely on the conversations around the feedback between students without structuring student activities and providing guidance during various feedback processes. Moreover, although dialogic peer feedback can be scalable, students' benefits from it can elevate with proper instructor facilitation (Zhu & Carless, 2018). Learning analytics (LA) can be used to support instructor facilitation by offering actionable insights toward student engagement during the feedback practice.

Attending to the given gap, this paper first presents a theoretical framework of dialogic peer feedback. Then, it introduces a web-based platform, called Synergy, designed (based on the framework) to facilitate dialogic peer feedback. LA is integrated into Synergy to enhance instructors' capacity to intervene timely and properly in situations when students need assistance.

2 THEORETICAL FRAMEWORK

The proposed theoretical framework suggests that in a successful dialogic feedback activity, students need to (1) initially plan the feedback and coordinate their activities, (2) then engage in feedback provision and discussion, and (3) finally take actions to translate the feedback into progress on the work at hand. This framework is underpinned by the Hadwin et al. (2011, 2017)'s work which theorizes that collaborative learning involves three types of regulated learning: socially shared regulation of learning (i.e., during feedback planning in the first phase), co-regulation of learning (i.e., during feedback discussion in the second phase), and self-regulation (i.e., when taking actions based on feedback). Within each of these phases several iterations might be necessary.

3 THE SYNERGY PLATFORM

The Synergy platform, grounded in a theoretical framework, provides a structured environment and rich set of tools to facilitate dialogic peer feedback in a systematic way. Reviewing students are assigned two tasks: assess the peer's work and provide feedback, whereas students being reviewed need to perform three tasks: assess the own work, read and discuss the feedback, and revise the work. Within these tasks there exists several sub-tasks. The alignment of these tasks with the theoretical framework is given in Table 1.

Table 1: The alignment between the theoretical framework and the design of Synergy

Phases in the theoretical framework	Associated review tasks	Associated components of Synergy
1. Planning and coordination of feedback activities	Assess the peer's work, Assess the own work,	Feedback planner, Assessment tool, Discussion tool (for feedback tasks and assessment scores)
2. Discussion of feedback to support its uptake	Provide feedback, Read and Discuss the feedback	Embedded Google Docs support
3. Translation of feedback into action	Revise the work	Action planner, Discussion tool (for learning actions), Progress Tracking

Within the scope of this paper, assessment tool, feedback planner, and action planner are briefly described.

3.1 Assessment tool

Assessment tool (see Figure 1) is designed to enable both peer and self-assessment. This tool displays the submitted work using Google Docs and allows assessing the assigned work based on the rubric created by the instructor.

Once an assessment is performed by a student (for own work or a peer's work), assessment tool provides a comparison of the scores assigned by all students (i.e., students being reviewed and students reviewing), as seen in Figure 2. Assessment items scored differently are listed for discussion to facilitate negotiation of perspectives about the quality of the work.

The work to be assessed is shown below. Use the rubric (just under the document) for the assessment.

```

1  ... 1  ... 2  ... 3  ... 4  ... 5  ... 6  ... 7  ...
fullname = "Mike Jumper"
isSpace = False
name = ""
surname = ""

length = len(fullname)

name = fullname[0] + fullname[1] + fullname[2] + fullname[3]
surname = fullname[5] + fullname[6] + fullname[7] + fullname[8] +
fullname[9] + fullname[10]

```

Rubric for assessing the programming assignment #2.

1. ASSESSING

Show the Items

The code properly uses the loops to minimize hard-coding.

1 feedback tasks

The code properly uses functions to reduce repetition and complexity.

The code produces the desired outcome correctly.

The code is well-documented and explained with comments.

The code runs correctly without syntax and runtime errors.

Assess 

Assess 

Assess 

Assess 

Assess 

Hide

Save the scores

Figure 1: Assessing a work in assessment tool

In order to see the current assessment scores, please choose a category:

ASSESSING

All assessment scores assigned are provided below. Rows represent the students and columns represent the assessment items.

Anonymous_2	2	2	4	4	2
Lilias E.	2	4	4	5	1
	R2	R3	R4	R5	R1

Discussion

Assessment items scored differently are listed below. Discuss with your peers to understand each other's perspective and to reach an agreement.

R1: The code properly uses the loops to minimize hard-coding.

Anonymous_1: 2 Anonymous_3: 1

Discuss

R3: The code produces the desired outcome correctly.

Anonymous_1: 2 Anonymous_3: 4

Discuss

Figure 2: Comparison of assessment scores

3.2 Feedback Tasks and Feedback Planner

Synergy uses the concept of *feedback task* to help students (reviewing a student work) plan their feedback ahead of time. Feedback tasks serve as notes that students take to plan their feedback. By using Feedback Planner (see Figure 3), students can view all the feedback task, create a new one, or edit an existing one. They can also discuss them by clicking on the Discuss button.

+ New Task

Feedback Planner helps you organize ahead of time the feedback that you want to provide for your peers. You can create *feedback tasks* to plan what feedback you plan to provide for a specific assessment criteria.

⚠ Remember that feedback tasks are notes you take for yourself to plan the feedback. The actual feedback should be posted on the Google Doc based on your plan.

Feedback tasks are listed below. For each task, the responsible peer(s) is shown on the right handside.

#1: Helping improve the while loop 🗑 ✎

Lilias Elcom

Discuss
🔗 Associated rubric item: The code properly uses the loops to minimize hard-coding.

#2: Helping to fix the code about creating functions 🗑 ✎

Lilias Elcom
 Rebekah Chipps

Discuss
🔗 Associated rubric item: The code properly uses functions to reduce repetition and complexity.

Figure 3: Feedback planner

3.3 Learning Actions and Action Planner

Synergy uses the concept of *learning actions* to help students (receiving feedback on their work) plan revisions on their work based on all feedback received. Students can create actions when checking the peers' feedback on their work. The actions created are displayed on the same page, as seen in the following figure.

+ Learning Action
Until now, 4 actions in total. [\[Show/Hide\]](#)

i You can later manage them through Action Planner

- To improve the code [Complete before: 2019-11-06]
- Refine the code about the loop [Complete before: 2019-11-06]
- Change the function [Complete before: 2019-11-06]
- Update the code about the function [Complete before: 2019-11-08]

YOUR SUBMISSION FOR ASSIGNMENT #1: STRING OPERATIONS

🔗 [Open in new window](#)

File Edit View Insert Format Tools Add-ons Help

```

fullname = "Mike Jumper"
isSpace = False
name = ""
surname = ""

length = len(fullname)
          
```

Lilias Elcom
 12:59 PM Today
 Resolve

I think the declaration of the variables is problematic. You should add the variable type.

Figure 4: Creating learning actions while reading feedback

Furthermore, students can use the *Action Planner* to manage the actions (e.g., creating new actions, editing/deleting existing ones, and discussing them) (see Figure 5).

[+ Learning action](#)

Action list

Action	Difficulty	Deadline ⓘ	
Changing the code about splitting strings	2 <small>Low</small>	2019-12-05	Delete Discuss
Updating the function	1 <small>Low</small>	2019-11-10	Delete Discuss

Figure 5: Action planner

4 LEARNING ANALYTICS SUPPORT

Learning analytics support for instructors is integrated into Synergy in three parts: 1) assessments and feedback planning, 2) feedback provision and discussion, and 3) action progress and revisions, which map to the phases suggested by the theoretical framework grounding Synergy. Each part is composed of a LA dashboard to provide actionable insights for instructors. The design of the dashboards is informed by the theory and justified by the empirical research from the literature.

4.1 Assessment and Feedback Planning Dashboard

This dashboard provides a class overview of student activities regarding assessments and feedback planning. The goal is to enable instructors to identify issues regarding student activities in assessing the assigned works, discussing the assessment results and aligning the perspectives, and planning the feedback. The dashboard (see Figure 6) first provides an overview of student engagement and highlights about assessment items.

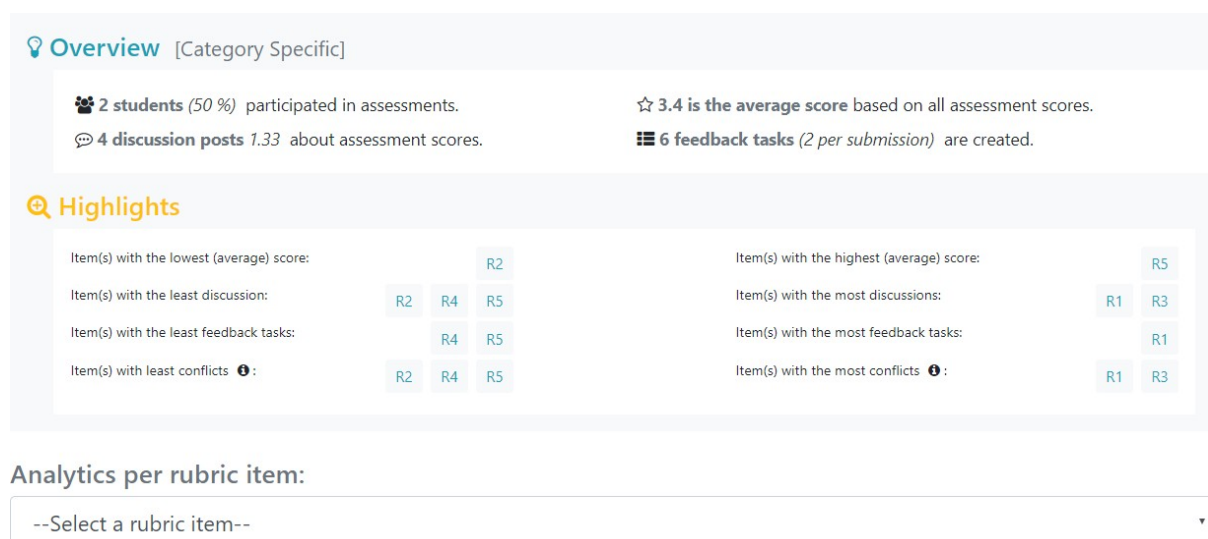


Figure 6: Overview and highlights in assessment and feedback planning dashboard

As seen in Figure 7, a specific assessment item can be selected to view details of student activities item-wise.

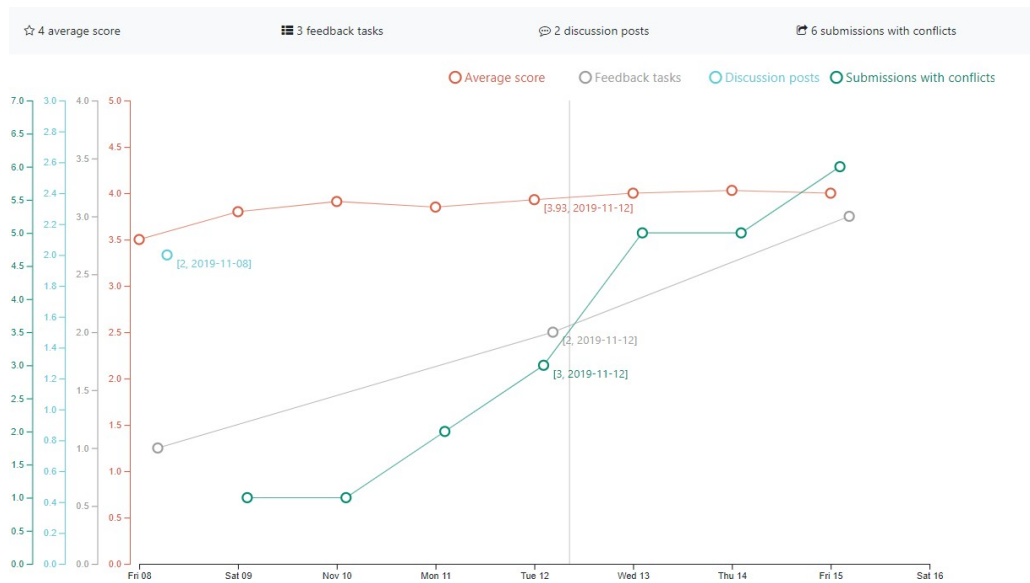


Figure 7: Scatter plot of assessment activities for a selected item

4.2 Feedback Provision and Discussion Dashboard

This dashboard aims to enable instructors to monitor the feedback provision and discussion activities of the whole class. The goal is to help instructors identify potential issues regarding students' engagement in providing feedback, discussing feedback, and deriving actions from feedback.



Figure 8: Feedback provision and discussion dashboard

4.3 Action Progress and Revisions Dashboard

Below (see Figure 9) is the dashboard to allow instructors monitor the overall progress on learning actions and efforts put in revising the works in whole class. The goal is to help instructors identify potential issues regarding low student engagement and progress.

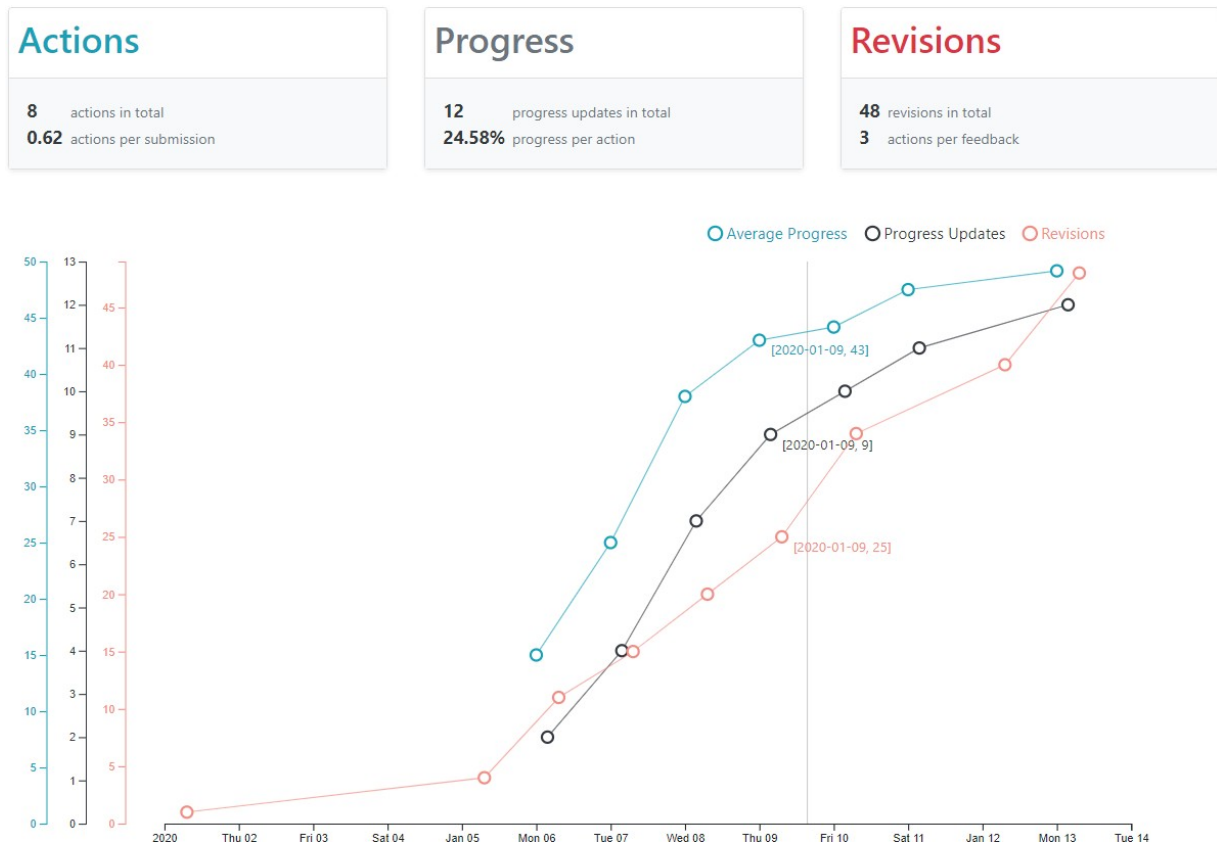


Figure 9: Action progress and revisions dashboard

REFERENCES

- Yang, M., & Carless, D. (2013). The feedback triangle and the enhancement of dialogic feedback processes. *Teaching in Higher Education*, 18(3), 285–297.
- Zhu, Q., & Carless, D. (2018). Dialogue within peer feedback processes: Clarification and negotiation of meaning. *Higher Education Research & Development*, 37(4), 883–897.
- Hadwin, A. F., Järvelä, S., & Miller, M. (2011). Self-regulated, co-regulated, and socially shared regulation of learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of Self-Regulation of Learning and Performance* (pp. 65–84).
- Hadwin, A. F., Järvelä, S., & Miller, M. (2017). Self-regulation, co-regulation and shared regulation in collaborative learning environments. In D. H. Schunk & J. Greene (Eds.), *Handbook of self-regulation of learning and performance* (2nd ed.). New York, NY: Routledge.

What matters?: The role of feedback in developing preservice teachers' digital competency for future teaching

Sarah K. Howard

University of Wollongong
sahoward@uow.edu.au

Jo Tondeur

Vrije Universiteit Brussel
Jo.Tondeur@vub.be

Jack Yang

University of Wollongong
jiej@uow.edu.au

ABSTRACT: Developing preservice teachers' digital competency to design technology-enhanced learning is a challenge. It requires integration of a range of strategies, one of which is Feedback. However, teacher educators often struggle to understand how their students experience learning to teach with technology, which limits how well they are able to provide appropriate feedback. Methods of analyzing (preservice) teachers' experiences and making it actionable as feedback are needed. The aim of this study is to create tools to reveal differences in preservice teachers' experiences developing digital competence, to support teacher trainer decision making about feedback to support this process. To this end, we reanalyzed preservice teacher questionnaire data ($N = 931$), looking at their reported experiences with digital competence strategies. Data is analyzed through clustering preservice teachers based on attitudes, and using a simple association rules approach with graph visualizations. Results reveal some of the complexity of developing digital competency in teacher training, by making key associations among Feedback and other strategies visible. Communicating these associations to teacher trainers, provides a first step in understanding appropriate ways to direct feedback in the development of digital competency and inform learning design. Implications for preservice teacher support will be explored.

Keywords: teacher feedback; digital competencies; teacher training; decision making; learning design

1 INTRODUCTION

To design learning experiences that integrate the digital skills and ways of working learners will need for the future, teachers will require a high level of digital competence. In order to achieve this in the teaching workforce, teacher trainers will need to purposively design learning experiences that will develop preservice teachers' digital competencies. However, this has proven difficult. One of the reasons for this difficulty is the complexity of developing digital competence (Mouza, Karchmer-Klein, Nandakumar, Ozden, & Hu, 2014), such as which strategies should be developed and how.

Effective strategies to develop digital competence in preservice teachers have been identified by Tondeur et al. (2012) in the Synthesis of Qualitative Data (SQD) model. The SQD-model outlines six effective strategies to develop preservice teachers' digital competence: Authentic experiences, Collaboration, Instructional design, Feedback, Reflection and Role models. However, they stressed that an approach was needed to unpack the complexity of relationships among strategies and make application in training clearer. Using data mining methods, able to handle complex relationships in data, Howard et al. (2020) conducted an initial analysis of preservice teachers' ($N = 931$) questionnaire responses about experiences with support and training in order to integrate technology into classroom activities.

In the current paper, we present a first step in extending from Howard et al.'s (2020) analysis, to explore these relationships in more depth. We begin with Feedback. The quality of feedback has been identified as one of the main problems with preservice teachers' developing digital competence (Tondeur et al., 2018; Mouza, et al., 2014). In the following we consider the relationship observed between Feedback and the other five SQD strategies, as experienced by preservice teachers. To do this, we first explore the nature of digital technology integration and present the SQD strategies, with a focus on unpacking the importance of Feedback. The use of data science approaches, specifically association rules, will be presented and explained. Results suggest strategies for how teacher trainers incorporate feedback in the development of digital competencies. How the work will be progressed will be discussed and implications for teacher training.

2 SUPPORTING COMPETENCIES FOR TECHNOLOGY INTEGRATION

2.1 Complexity

Digital technology integration has been described as a 'wicked problem' (e.g. Borko, Whitcomb & Liston, 2009; Lim, et al., 2013). Borko et al. (2009) define 'wicked problems' as "those that include a large number of complex variables – all of which are dynamic, contextually bound and interdependent" (p. 3). Key aspects of this complexity are continual change and digital technologies continue to change at a rapid rate, where individuals are constantly 'catching up' with new skills and ways of working. Moreover, teaching practice and digital technology use are rooted in educational contexts, which include a complex range of stakeholders with their own beliefs, values, preferences, etc. Preservice teachers cannot be taught how to use a type of digital technology, they must be equipped with competencies to be able to navigate future technology integration. This includes critical thinking about digital technologies, matching technology affordances with learning needs, and understanding students' experiences with digital technologies -- just to identify a few points. Designing teacher training to include this kind of learning is not straightforward, it is largely experiential, and presents its own questions and levels of complexity.

2.2 SQD and Feedback

In response to the need to develop teachers' digital competency, Tondeur et al. (2012) developed the SQD model. This comprises six key strategies that need to be in place in teacher training programs to develop digital competency (see Table 1).

The SQD strategies provide the kinds of experiences needed in teacher training to develop capabilities to engage with digital technology integration in sophisticated ways, and to design rich learning environments. In the current study, we specifically focus on the Feedback strategy and its role in preservice teachers' experiences. The pre-service teachers' qualitative comments in the Tondeur (2012) review indicate that on-going and process-oriented feedback of experts were beneficial to building their abilities to use technology in the classroom. At the same time, it seems that providing pre-service teachers adequate feedback can be considered challenging for teacher training institutions (see Mouza et al., 2014). Preparing preservice teachers to use digital technologies in teaching cannot be planned independently from other strategies (Mouza et al., 2014). To illustrate, feedback can be beneficial during the design process (DE), the authentic experiences (AU) or even during the collaboration. Nevertheless, little is known about where preservice teachers experience feedback in the training process or find it useful. This brings us to the main aim of the study.

Table 1: The six SQD strategies for digital competence

Role models (RO)	Providing examples and is a crucial motivator for the development of digital competencies
Reflection (RE)	Discussing and reflecting about the opportunities and risks of digital technology use in education
Instructional design (DE)	Providing opportunities to learn about technology integration through design
Collaboration (CO)	Mitigates feelings of insecurity when preservice teachers need to design ICT related curriculum
Authentic experiences (AU)	Experience the value to use ICT in education in authentic settings, doing rather than watching
Feedback (FE)	Feedback should be continually provided through discussions, questionnaires, interviews, and observations in order to follow how ICT competence develops, and what kind of problems preservice teachers face in using ICT.

3 APPROACH AND METHOD

To be able to explore the complexity of developing digital competence strategies, we have drawn on data mining approaches, specifically association rules analysis. Rittel and Webber (1973) highlight, among other things, that wicked problems cannot be analyzed to ‘expected outcomes’, that there is no solution to the problem, but that any solution attempted will change the system. Data mining, as an approach, is able to handle a large number of interrelated factors, but importantly, the approach does not presuppose a linear solution or match results against a linear model. Thus, it does not presuppose a solution or expect particular behaviors. This provides a greater likelihood that natural self-organization and emergence of the system can be observed.

The aim of the current analysis is to specifically explore the role of feedback as a digital competency strategy in teacher training. In Howard et al.’s (2020) recent analysis, questionnaire responses on preservice teachers’ (N = 931) experiences with the six SQD strategies, across 20 Belgian universities, well explored. The questionnaire was a self-report instrument to measure pre-service teachers’ perceptions of the extent to which they experience the necessary support and training in order to integrate technology into classroom activities. It was completed near the end of their training. The SQD scale included 20 items, using a five-point Likert-type scale. Data from the SQD scale were analyzed using association rules analysis (apriori algorithm). A 20-rule solution was initially selected for preliminary exploration of relationships among strategies. Rules were identified based on having over .80 confidence and .15 support. Results were then visualized using a directed graph. Building on Howard et al.’s (2020) initial work, in the following discussion we focus on the role of Feedback.

4 RESULTS

The directional graph represents the two 20 rules in the analysis of questionnaire data (see Figure 1). In the current analysis, we focus on Clusters 1 and 2. For a more in depth analysis please see Howard et al.’s (2020) initial analysis of this graph. Results show that if preservice teachers felt they were receiving quality Feedback (light green, Fe3) they were also likely to feel positively about Instructional Design (De3, light blue square) and Authentic Experiences (Au3, dark green square). This suggests, for this group, feedback was important in relation to these two strategies. Significant relationships between

Feedback and Role Models (Ro, pink), Reflection (Re, orange) and Collaboration (Co, brown) were not observed in this analysis.

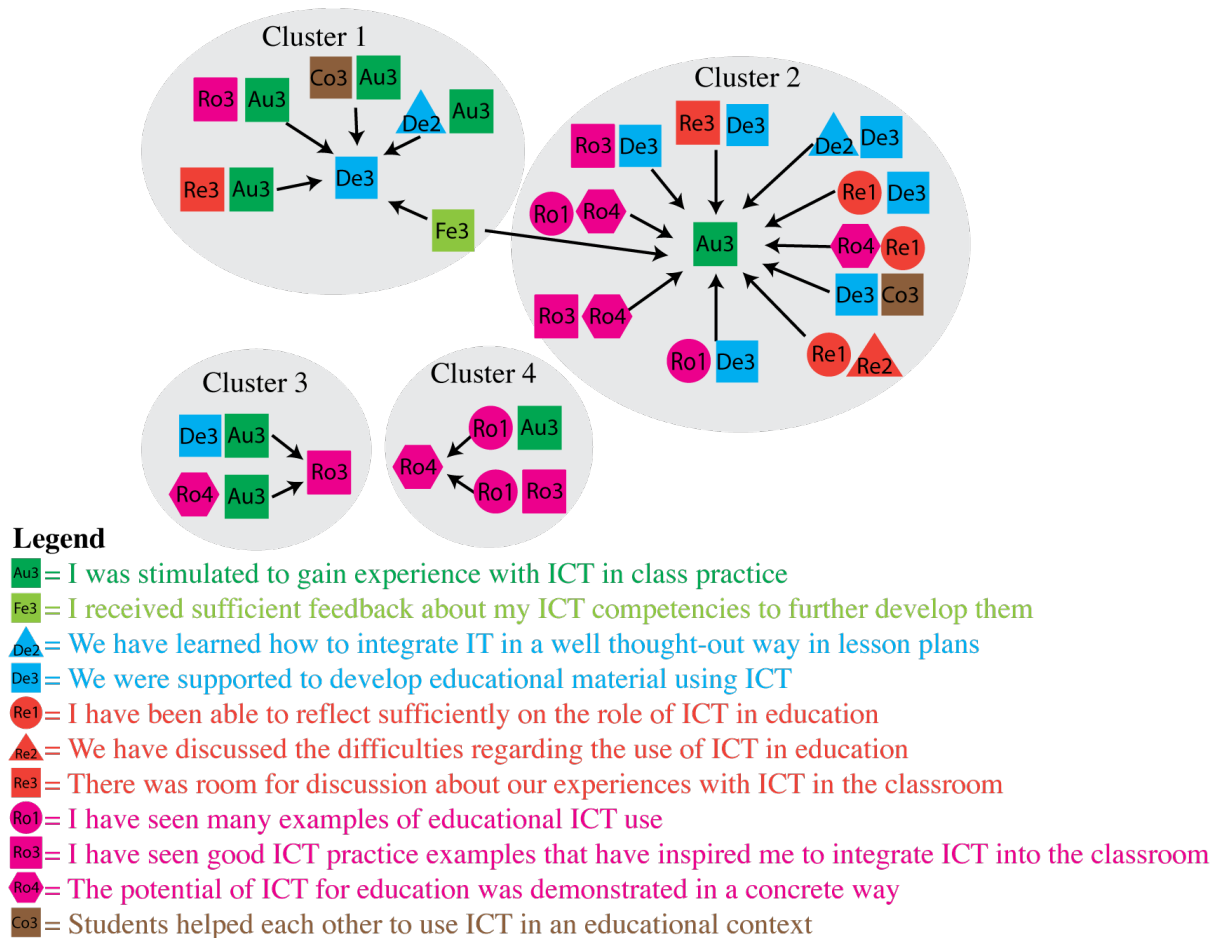


Figure 1. Directional graph of the 20-rule solution

5 DISCUSSION AND CONCLUSIONS

This analysis can guide teacher trainers to understand how preservice teachers are experiencing feedback. Here, positive experiences with feedback were most frequently occurring in relation to Design and Authentic Experiences. This finding provides two key insights. First, the association between feedback and Design, and Authentic Experiences, suggests that feedback on the other three strategies was either more mixed or less frequently occurring. This can inform how teacher trainers incorporate future feedback points, such as taking new approaches to supplying feedback in the other three areas (e.g. Kleinknecht & Gröschner, 2016). The second consideration may be around why these two strategies were related with Feedback. It could be argued that Design and Authentic Experiences closely relate to actual practicalities of integrating digital technologies (e.g. Stahl, Sharplin, & Kehrwald, 2016), while Reflection, Role Models and Collaboration support preservice teachers' understanding of that process. Feedback may have a more direct relationship to Design and Authentic Experiences, so it may be more appropriate to focus feedback here.

While Figure 1 presents an initial exploration of this area, it is an unrefined model. The next step will be to further explore the data. Given the six strategies in the SQD-model, this analysis provides a way to narrow design decisions about where feedback can be integrated with the two strategies. However,

this is a large group and the role of feedback may change for sub-groups, such as institutions, cohorts and preservice teachers with different attributes, e.g. attitudes about digital technologies. Moreover, individuals may have unique experiences. The aim would be to ultimately devise a classification model able to clarify some of the complexity in where feedback is best integrated for different groups of students or even individuals. This may employ methods such as Classifications Based on Associations, which can support teacher trainers' decisions about which strategies feedback may be best focused, based on other 'similar' preservice teachers.

Future implications of this work will most significantly affect teacher training design. Specifically, a tool such can provide a mechanism to support teacher trainers to responsive to preservice teachers' heterogeneous experiences developing digital competence. This is essential, given the complexity of developing digital competence for future teaching. Feedback is a critical component of this process, particularly considering the level of critical engagement and design necessary to understand digital technology integration. Without a better understanding of their different experiences and what matters to different groups, how to fully prepare future teachers for technology integration will remain obscured.

REFERENCES

- Howard, S. K., Tondeur, J., Ma, J., & Yang, J. (2020). Seeing the wood for the trees: Insights into the complexity of developing pre-service teachers' digital competencies for future teaching. In ASCILITE 2019: 36th International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education.
- Howard, S. K., Ma, J., & Yang, J. (2016). Student rules: Exploring patterns of students' computer-efficacy and engagement with digital technologies in learning. *Computers & Education*, 101, 29–42. <https://doi.org/10.1016/j.compedu.2016.05.008>
- Kleinknecht, M., & Gröschner, A. (2016). Fostering preservice teachers' noticing with structured video feedback: Results of an online- and video-based intervention study. *Teaching and Teacher Education*, 59, 45–56. <https://doi.org/https://doi.org/10.1016/j.tate.2016.05.020>
- Mouza, C., Karchmer-Klein, R., Nandakumar, R., Ozden, S. Y., & Hu, L. (2014). Investigating the impact of an integrated approach to the development of preservice teachers' technological pedagogical content knowledge (TPACK). *Computers & Education*, 71, 206–221.
- Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a General Theory of Planning. *Policy Sciences*, 4(2), 155–169. <https://doi.org/10.2307/4531523>
- Stahl, G., Sharplin, E., & Kehrwald, B. (2016). Developing pre-service teachers' confidence: real-time coaching in teacher education. *Reflective Practice*, 17(6), 724–738. <https://doi.org/10.1080/14623943.2016.1206882>
- Tondeur, J., Aesaert, K., Prestridge, S., & Consuegra, E. (2018). A multilevel analysis of what matters in the training of pre-service teacher's ICT competencies. *Computers & Education*, 122, 32–42. <https://doi.org/10.1016/j.compedu.2018.03.002>
- Tondeur, J., van Braak, J., Sang, G., Voogt, J., Fisser, P., & Ottenbreit-Leftwich, A. (2012). Preparing pre-service teachers to integrate technology in education: A synthesis of qualitative evidence. *Computers & Education*, 59(1), 134–144. <https://doi.org/10.1016/J.COMPEDU.2011.10.009>

Integrating Multi-channel Learning Data to Model Complex Learning Processes

Roger Azevedo

University of Central Florida
roger.azevedo@ucf.edu

George Siemens

University of Texas, Arlington and the University of South Australia
George.Siemens@unisa.edu.au

Shane Dawson

University of South Australia
Shane.Dawson@unisa.edu.au

ABSTRACT: During this full-day workshop on integrating multi-channel learning data to model complex learning processes, we include participants collaboratively articulating models of self-regulated learning (SRL) for any existing learning system (e.g., MetaTutor, MOOCs, etc.) to discuss issues about collecting and analyzing multimodal multichannel data and their implications for learning analytics. Particular focus will be on the technology environment required to integrate multimodal data challenges such as time stamping data that has different sources (e.g., logfiles and psychophysiological) related to learning processes that can only be evaluated at different rates (e.g., boredom has a longer psychological window of measurement than surprise does) and matching those data sources to theoretically-based constructs. Follow-up discussions will examine learning analytics, self-regulated learning processes, and ways to design data visualizations to model, foster, and support learners' self-regulated learning. During this workshop, we plan to discuss measurement and pedagogical issues related to making existing learning systems more intelligent (e.g., embedding data visualizations capable of scaffolding learning based on learning analytics).

Keywords: learning processes, self-regulated learning, multimodal data, methodologies, data analytics, data visualizations.

1 WORKSHOP BACKGROUND

Learning with advanced learning technologies (ALTs) such as intelligent tutoring systems, serious games, simulations, immersive virtual learning environments, and MOOCs, involves intricate and complex interactions between cognitive, metacognitive, motivational, affective, and social processes across different tasks and contexts. Current psychological and educational research on learning with ALTs provide a wealth of empirical data indicating that learners of all ages have difficulty learning about complex topics in areas such as STEM across contexts. Learning with ALTs requires students to analyze the learning situation, set meaningful goals, determine which strategies to use, assess whether those strategies are effective for meeting their goals, and then evaluate their emerging understanding of the topic. After evaluating, students can choose whether to modify their plans, goals, strategies, and efforts in relation to contextual conditions (e.g., cognitive, motivational, resources, and task conditions). Further, depending on the learning task, they need to continuously reflect on their learning. A major challenge for researchers, educators, instructional designers,

learning engineers, and data scientists involves collecting, tracking, modeling a myriad of complex processes using a variety of methods, tools, and sensors (e.g., synchronizing time, matching trace data to cognitive processes, making instructional decisions to optimize learning). We argue that understanding the role of these processes requires measuring, analyzing, and modeling multimodal multichannel data (e.g., logfiles, eye tracking, physiological sensors, facial expressions of emotions) during learning and problem solving with ALTs across formal (e.g., school) and real-world contexts (e.g., online learning, military, industry, informal learning). Understanding the complex nature of unfolding SRL processes has recently been addressed by emerging interdisciplinary research using online trace methods (e.g., logfiles, eye tracking, think-aloud protocols, physiological sensors, screen recording of human-machine interactions, classroom discourse).

Using these methods has been widely applauded by the research community due to the advantages over traditional methodologies (e.g., self-report measures) which provide evidence regarding: (1) augmenting the descriptive and explanatory adequacy of current models of cognition, learning, instruction, and SRL; (2) capturing real-time unfolding processes within context; (3) understanding how internal (e.g., prior knowledge) and external (e.g., level of external regulation by human or artificial agent) factors and other variables (e.g., individual differences) impact the use and quality of these processes (e.g., negative affective reaction to an avatar's prompting to use a cognitive strategy) during learning; (4) generating predictions and hypotheses about the interactions between specific processes (e.g., relationship between metacognitive monitoring accuracy and emotion regulation strategies); (5) measuring the quantity and quality of these processes on embedded assessments (e.g., quizzes, summaries), instructional choices (e.g., compliance with external regulation by human and artificial agents, persistence, self-efficacy), learning outcomes, transfer, and so forth; and (6) making real-time adaptations based on analyzing data using data mining and machine-learning techniques.

Despite the benefits of using multimodal multichannel data, it comes with its own set of challenges to be addressed by the participants of this workshop. They include the following: (1) temporally aligning data sources based on different sampling rates; (2) the “right” amount of data needed to be sampled in order to accurately classify and infer the underlying processes; (3) level of granularity at which the classified and inferred data are made (e.g., macro, micro, or valence level; duration of an affective state; psychological and educational meaning of a physiological event); (4) impact of data transformations from raw data to actionable data using dashboards and other learning analytics approaches; (5) complexity in dealing with noisy and messy data (e.g., missing data) with traditional and contemporary data mining and machine-learning techniques; (6) embodying theoretical assumptions in data streams (e.g., three revisits from eye gaze behavior data from at least two different areas of interest constitute a monitoring process of content evaluation) related to classifying and inferring; (7) assessing the level of accuracy, modeling (the human and machine) complex underlying processes, and confidence in inferring based on current analytical methods; (8) ascertaining the correct level of “classifying” depending on the intended use of the multichannel data and implementing it into ALT architectures; and (9) the analytical bottleneck created when converging single and multichannel data and latency in using generated inferences for instruction and learning (e.g., increased latency might miss opportunity to provide timely scaffolding needed to facilitate emotion regulation) and possible consequences (e.g., delayed adaptive scaffold might lead to negative affective reaction). **Lastly, these and other questions will focus on the conference's theme—i.e., what are the implications and potential impact of the presented work for the next 10 years (e.g., what are the practical and scholarly implications of the presented work for the next ten years? What are the challenges of the presented work we need to address to improve its impact in the next ten years? How can the presented work be practically implemented and adopted?).** These are some of the issues currently being addressed by interdisciplinary researchers that will be targeted by our workshop.

1.1 Organizational Details of the Workshop

Type of event: Workshop // Proposed schedule and duration: Half-day

Type of participation: mixed participation including: (1) invited interdisciplinary researchers; (2) those who submit papers, posters, and demos to the workshop; (3) and open participation workshop to anyone interested may register to attend.

Expected participant numbers and planned dissemination activities to recruit attendants: approximately 40 participants and recruit through workshops website, emails to vast networks so international researchers belonging to the LAK community as well as other communities (e.g., American Educational Research Association [AERA], European Association for Research on Learning and Instruction [EARLI], Society for Artificial Intelligence in Education [AIED], Intelligent Tutoring Systems [ITS], International Society of the Learning Sciences [ISLS], Cognitive Science, etc.), special invitations to invited presenters, etc.

1.2 Workshop Format and Planned Activities

During the workshop, we will have presentations focused on workshop topics with ample time for discussion. Participants will collaboratively articulate models of self-regulated learning (SRL) for any existing learning system (e.g., MetaTutor, BioWorld, Betty's Brain, Crystal Island, MOOCs, immersive virtual environments, etc.) as well as discuss the issues involved in collecting and analyzing multimodal multichannel data and their implications for learning analytics. Particular focus will be on the technology environment required to integrate multimodal data (e.g., challenges such as time stamping data from different sources (e.g., logfiles and psychophysiological sensors) related to psychological processes that can only be evaluated at different rates (e.g., boredom has a longer psychological window of measurement than surprise does in relation to their influence on learning) and matching those data sources to specific theoretically-based constructs (e.g., cognitive load, self-efficacy, motivation, cognition, metacognition, etc.). Follow-up discussion will be on learning analytics, self-regulated learning processes, and designing data visualizations to model, foster, and support learners' self-regulated learning. In addition, we will have small group sessions to brainstorm about novel methodologies and analytical techniques used to evaluate cognitive, affective, metacognitive, motivational, and social processes during human-machine interactions and their implications for learning analytics and data visualizations. Further, we will have interactive demos of existing systems and other prototypes, especially on the methods used in collecting multimodal multichannel SRL data as well as generating data visualizations and the challenges and advantages that each bring. We also plan to instrument one of the graduate students or workshop participants and analyze their real-time multimodal multichannel data as they use an existing ALT (e.g., MetaTutor) and discuss implications for data analytics and data visualizations. During this activity, we plan to discuss measurement and pedagogical issues related to making ALTs more intelligent (e.g., embedding data visualization capable of scaffolding learning based in learning analytics). Finally, we will also set aside specific time for discussing opportunities and pathways for cross-institutional and industry collaborations.

1.3 Workshop Objectives and Intended Outcomes

1. Unite interdisciplinary researchers to share, explain, and discuss conceptual, theoretical, methodological, analytical issues related to multimodal multichannel data and learning analytics.
2. Share advanced statistical, data mining, and machine-learning methods for analyzing complex, multimodal multichannel process data and discuss implications for education and training via data visualizations (e.g., dashboards, intelligent virtual humans, etc.).
3. Present and discuss strengths and weaknesses associated with collecting multimodal data, coding schemes, data pipelines, algorithms, synchronizing and transforming data, etc. to create and extend an international network that will allow for sharing of resources between researchers across disciplines and locations.

4. Submit the results of workshop as a LAK Companion proceedings, special issue on interdisciplinary journal (e.g., *Journal of Learning Analytics*, *Computers in Human Behavior*, *IEEE Transactions on Learning Technologies*, *Journal of the Learning Sciences*, *Learning & Instruction*, etc.) that will lead to publications and symposia at several international conferences (e.g., AIED 2021, Cognitive Science 2021, and EARLI 2021).
5. Submit team-based grants to several NSF programs (e.g., Cyberlearning and Future of Work, DRK12, Advancing Informal Stem Learning, Science of Technology Center, Future of Work-Human Technology Frontier) as well as other funding agencies and foundations.
6. Develop new partnerships with industry (e.g., Amazon, Boeing, Southwest Airlines, Walt Disney) and other government institutions and private organizations to pursue 4. and 5.

1.4 Structure and contents of the workshop website

Please refer to our website (<https://sites.google.com/view/multi-channellearning/home>) that advertises our workshop and include key information, including (1) organizers and their contacts info; (2) access/link to LAK 2020 conference website (<https://lak20.solaresearch.org/>); (3) information about the workshop (e.g., important dates, structure, submission guidelines, submission types, etc.); (4) list of workshop contributors and their papers and slides (with permission); and (5) associated with (4), and in a protect section of the website, authors may include sample data (e.g., video clips of facial expressions, screen recordings, eye tracking data, logfiles, etc.) as well as coding schemes, research protocols, etc.

1.5 Key References

- AZEVEDO, R., & GAŠEVIĆ, D. (2019). ANALYZING MULTIMODAL MULTICHANNEL DATA ABOUT SELF-REGULATED LEARNING WITH ADVANCED LEARNING TECHNOLOGIES: ISSUES AND CHALLENGES. *COMPUTERS IN HUMAN BEHAVIOR*, 96, 207-210.
- BANNERT, M., REIMANN, P., & SONNENBERG, C. (2014). PROCESS MINING TECHNIQUES FOR ANALYZING PATTERNS AND STRATEGIES IN STUDENTS' SELF-REGULATED LEARNING. *METACOGNITION & LEARNING*, 9, 161-185.
- DAWSON, S., JOKSIMOVIC, S., POQUET, O., & SIEMENS, G. (2019). INCREASING THE IMPACT OF LEARNING ANALYTICS. *PROCEEDINGS OF THE ANNUAL MEETING OF THE LEARNING ANALYTICS AND KNOWLEDGE CONFERENCE (LAK 2019)*.
- DUNLOSKY, J., & RAWSON, K. (2019). *THE CAMBRIDGE HANDBOOK OF COGNITION AND EDUCATION*. CAMBRIDGE, MA: CAMBRIDGE UNIVERSITY PRESS.
- JÄRVELÄ, S., JÄRVENOJUA, H., & MALMBERG, J. (2019) CAPTURING THE DYNAMIC AND CYCLICAL NATURE OF REGULATION METHODOLOGICAL PROGRESS IN UNDERSTANDING SOCIALLY SHARED REGULATION IN LEARNING. *INTERNATIONAL JOURNAL OF COMPUTER-SUPPORTED COLLABORATIVE LEARNING*, 14, 425-441.
- KINNEBREW, J. S., SEGEDY, J. R., & BISWAS, G. (2015). INTEGRATING MODEL-DRIVEN AND DATA-DRIVEN TECHNIQUES FOR ANALYZING LEARNING BEHAVIORS IN OPEN-ENDED LEARNING ENVIRONMENTS. *IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES*, 10, 140-153.
- LANG, C., SIEMENS, G., WISE, A., & GASEVIC, D. (2017). *HANDBOOK OF LEARNING ANALYTICS*. VANCOUVER, BC: SOCIETY FOR LEARNING ANALYTICS RESEARCH.
- SIEMENS, G. (2019). LEARNING ANALYTICS AND OPEN, FLEXIBLE, AND DISTANCE LEARNING. *DISTANCE EDUCATION*, 40, 414-418.
- WINNE, P. H. (2018). COGNITION AND METACOGNITION WITHIN SELF-REGULATED LEARNING. IN DALE H. SCHUNK & JEFFREY A. GREENE (EDS.), *HANDBOOK OF SELF-REGULATION OF LEARNING AND PERFORMANCE* (PP. 52-64). NEW YORK: ROUTLEDGE.

Contextualizing multimodal learning analytics to theoretical frameworks and learning environments

Elizabeth B. Cloude

University of Central Florida
elizabeth.cloude@knights.ucf.edu

Roger Azevedo

University of Central Florida
roger.azevedo@ucf.edu

ABSTRACT: Technology easily captures millions of data points from learning environments (e.g., virtual learning environments [VLEs]) which are used to derive insights that predict and improve performance. Yet, challenges remain because most studies use data-driven approaches to build models of learning without accounting for the context in which learning occurred such as accounting for the resources and constraints that different learning environments offer (e.g., note taking and summarizing tools vs no tools). In this presentation, we propose analytical techniques which address this challenge by accounting for theoretical frameworks and empirical findings during learning. Specifically, we provide an example of this method using thirty-seven learners' ($n = 37$) multimodal data which were collected during learning with a VLE in a classroom. Data were aligned with the information processing theory of self-regulation where variables were mapped onto theoretically-based constructs of learning. Implications of this research could provide a means to collect, process, and analyze multimodal data that account for the contextual resources and constraints learners may face to inform and promote effective personalized scaffolding and feedback that optimize learning with technology.

Keywords: contextualized multimodal data, virtual learning environments, self-regulated learning, personalized scaffolding and feedback

1 BACKGROUND

Educational and learning scientists investigate factors related to learning outcomes such as the processes and strategies involved during learning and their role in performance. Multimodal data capture what learners do over a learning session using sensors and devices such as an electrodermal bracelet (Lane & D'Mello, 2019), on-line trace data (Azevedo et al., 2013), or emotion detection software (Taub et al., in press). To understand these data, previous studies have employed data-driven methods to develop predictive models of learning. However, this analytical approach assumes a general model of learning for all learners regardless of individual differences, self-regulated learning competency, and the constraints that various environments may impose on learning outcomes. Since empirical literature suggests contextual factors impact learning over time such as the learning technology (e.g., serious game), domain (e.g., art vs. math), and setting (e.g., classroom vs. laboratory; Azevedo & Gašević, 2019; Mangaroska & Giannakos, 2018; Matcha, Gašević, & Pardo, 2019; Shibani, Knight, & Shum, 2019; Roll & Winne, 2015), we argue data-driven methods miss critical data answering *why* a learner initiated a learning process (or lack thereof), explaining learning outcomes and performance.

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

Virtual learning environments (VLEs) are an important medium to consider for multimodal learning analytics (MLA) since VLEs immerse learners in an entirely new context, where the system presents a vivid and rich world shutting off physical reality and generating feelings of “being there” during learning. The system’s immersive features can be controlled or manipulated such as placing learners in a microscopic plant cell vs an animal cell. Data collected during learning with VLEs provide researchers with opportunities to investigate *how*, *when*, *why*, and *what* contextual factors impact learning and performance by changing the world in which learning occurs (Cummings & Bailenson, 2016; Lee, Wong, & Fung, 2010). As such, we ask a fundamental question. *How can multimodal data be contextualized to the learner and their learning session?* We contextualized a dataset of 37 learners’ multimodal data during learning with a VLE to answer this question.

1.1 Contextualizing multimodal learning data

Self-regulated learning (SRL) is defined as a learner continuously monitoring and controlling their cognitive, metacognitive, affective, and motivational processes to achieve an objective (Azevedo et al., 2017). We used the information processing theory of SRL (Winne & Hadwin, 1998) to operationalize data since this theoretical framework accounts for context during learning. Specifically, information processing theory of SRL explains a model that highlights five facets occurring within and across phases of SRL: (1) conditions which represent the resources and constraints presented in a task, tapping into (a) internal—learner’s cognitive ability such as prior knowledge and SRL competency, and (b) external—resources and constraints in the environment such as limited time on task, (2) operations, or specific strategies used during learning, (3) products such as the knowledge gained from (2), (4) evaluations, or examining how well knowledge gained in (3) contribute to meeting the objective, and (5) standards, the criterion with which (3) products are evaluated against in (4) (e.g., content assessment, beliefs; COPES; Winne, 2017). To clearly demonstrate how to map multimodal data to theoretical frameworks and account for context, we used 37 learners’ multimodal data captured during learning about photosynthesis with a VLE and provide examples of ways to map and operationally define variables aligned with information processes theory of SRL and COPES model.

2 MAPPING DATA TO THEORETICALLY BASED LEARNING CONSTRUCTS

Thirty-seven ($n = 37$) high schoolers learned with a VLE and we captured (1) pre/post-test self-reported presence, motivation, emotions and values, and self-efficacy using questionnaires before and after learning with a VLE; (2) pre/post-test scores using an 11-item, multiple-choice assessment on photosynthesis before and after learning with a VLE; and (3) real-time process data during learning with a VLE using concurrent verbalizations. To operationalize the external factors involved in conditions, it is critical to explain the nature of the VLE. The VLE was designed to teach photosynthesis concepts, with the overall objective of requiring learning to generate as many glucose molecules as possible during two phases: (I) light-dependent reactions and (II) light-independent reactions. Both phases I and II had restricted time frames, imposing constraints on learning about photosynthesis to 2.5 minutes each. Specifically, phase I required learners to shoot photons through thylakoids which triggered a splitting water mechanism. This task generated hydrogen and oxygen ions, required to complete phase II and presenting another constraint in task I. To succeed in phase II, learners needed to shoot photons through at least six thylakoids to generate enough hydrogen ions to generate one molecule of glucose. This presented another a potential

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

constraint or resource in task II, such that if the learner did not generate at least six hydrogen ions, then they would not be successful in completing task II, whereas if the learner generated many hydrogen ions, they would have enough resources to assemble multiple glucose molecules. Accounting for these external factors in phases I and II are critical for understanding *why* a learner may be initiating a learning process or strategy. The constraints and resources may also inform other COPES processes such as operations used. Additionally, learners' internal conditions such as prior knowledge may impact their ability to adjust to the demands (i.e., constraints) and affordances (i.e., resources) of a learning environment. For instance, the learner may not understand that photons must go through thylakoids as opposed to chloroplasts to split water molecules, impeding their time available to generate hydrogen ions. As such, all of these contextual factors need to be accounted for when variables are processed and analyzed to understand true nature of learning and SRL processes. We propose a method (see Table 1) that contextualizes learning with a VLE by accounting for COPES, a critical part of self-regulation.

Table 1: Contextualizing multimodal data to information processing SRL theory and COPES model

Theoretical constructs	Modality	Variable
Conditions	1. Pre-test scores/self-reported self-efficacy	1. Number of correct answers/scores of
1. Internal		
2. External	2. VLE tasks I & II	2. Successfully complete task I or II
Operations	Concurrent verbalizations	Frequency of cognitive strategies
Products	Post-test scores	Number of correct answers
Evaluations	Concurrent verbalizations	Frequency of metacognitive processes
Standards	Concurrent verbalizations/self-reported motivation, emotions and value, self-efficacy	Frequency of interest/affect verbalizations/self-reported scores on questionnaires

3 FUTURE DIRECTIONS

Since Winne and colleagues SRL framework and COPES model (Winne & Hadwin, 2008; Winne, 2017) outline and describe a non-linear and dependent nature of learning processes such that internal conditions may be related to other facets of COPES, it is critical to capture contextual information to inform teaching and learning practices. However, future studies adopting this methodological and analytical approach need to first consider how different learning technologies (e.g., serious games, VLEs) affect learning processes. Another consideration for researchers should be around the best analytical technique that can handle the non-linear and dependent nature of learning variables that are informed by the theoretical framework used. For example, traditional inferential statistics such as multiple linear regression need to be approached cautiously, such that all variables reflecting COPES cannot be included as predictor variables in the same model to explain performance. Since COPES variables are theoretically related to one another, it begs the question of whether these relations should be statistically accounted for. Instances of multicollinearity have shown to inflate regression coefficients, producing more random noise and error in the model. Implications of this

method could lead to contextualized MLA which promote effective and personalized scaffolding and feedback based on individual learning needs and the environment in which learning occurred.

REFERENCES

- Azevedo, R., & Gašević, D. (2019). Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges. *Computers in Human Behavior*, 96, 207-210.
- Azevedo, R., Harley, J., Trevors, G., Duffy, M., Feyzi-Behnagh, R., Bouchet, F., & Landis, R. (2013). Using trace data to examine the complex roles of cognitive, metacognitive, and emotional self-regulatory processes during learning with multi-agent systems. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 427-449). New York, NY: Springer.
- Azevedo, R., Taub, M., & Mudrick, N. V. (2017). Understanding and Reasoning about Real-Time Cognitive, Affective, and Metacognitive Processes to Foster Self-Regulation with Advanced Learning Technologies. In J. A. Greene & D. H. Schunk (Eds.), *Handbook on self-regulation of learning and performance* (2nd ed, pp. 254–270). New York, NY: Routledge.
- Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, 19, 272-309.
- Lane, H. C., & D'Mello, S. K. (2019). Uses of physiological monitoring in intelligent learning environments: A review of research, evidence, and technologies. In T. Parsons, L. Lin, D. Cockerham (Eds.) *Mind, Brain and Technology. Educational Communications and Technology: Issues and Innovations* (pp. 67-86). New York, NY: Springer.
- Lee, E. A. L., Wong, K. W., & Fung, C. C. (2010). How does desktop virtual reality enhance learning outcomes? A structural equation modeling approach. *Computers & Education*, 55, 1424-1442.
- Mangaroska, K., & Giannakos, M. N. (2018). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies*.
- Matcha, W., Gasevic, D., Pardo, A., & others. (in press). A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. *IEEE Transactions on Learning Technologies*. doi.org/10.1109/TLT.2019.2916802
- Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, 2, 7-12. doi.org/10.18608/jla.2015.21.2
- Shibani, A., Knight, S., & Shum, S. B. (2019). Contextualizable learning analytics design: A generic model and writing analytics evaluations. Proceedings from LAK'19: *The 9th International Conference on Learning Analytics & Knowledge* (210-219). New York, NY: ACM. doi.org/10.1145/3303772.3303785
- Winne, P. H. (2017). Cognition and metacognition within self-regulated learning. In J. A. Greene & D. H. Schunk (Eds.), *Handbook on self-regulation of learning and performance* (2nd ed, pp. 52-64). New York, NY: Routledge.
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ: Lawrence Erlbaum Associates.

Measuring Micro-Level Self-Regulated Learning Processes with Enhanced Log Data and Eye Tracking Data

Fan¹, Y., Lim², K.P., van der Graaf³, J., Kilgour¹, J., Engelmann², K., Bannert², M.,
Molenaar³, I., Moore¹, J., Gasevic¹, D.

University of Edinburgh¹

Technical University of Munich²

Radboud University³

yizhou.fan@ed.ac.uk

ABSTRACT: To advance our understanding of processes that learners engage in self-regulated learning (SRL), we need novel approaches to measurement and integration of multi-channel data. Learning analytics has been recognized as a field that can offer unobtrusive measures of SRL processes through the use of log data. However, log data are insufficiently to capture the full scope of SRL processes. In this paper, we present the preliminary findings of a study that aimed to explore the extent to which the integration of eye-tracking data with log-data can advance detection of SRL processes such as orientation, planning and monitoring, as theorized about SRL in the literature. For detection of SRL processes in this combined eye-tracking and log data, a special library of action patterns was developed. Our results show that the joint eye tracking data and log data provided richer information about the learning areas of interest, and thus, greatly improved the granularity of measurement of SRL processes. In order to further validate the value of joining eye-tracking and log data, the future work will include the use of think-aloud data.

Keywords: self-regulated learning; enhanced trace data; eye track data; learning analytics;

1 INTRODUCTION

Self-regulated learners use cognitive processes (e.g., read, code and elaborate) to study a topic, engage in metacognitive activities (e.g., plan, monitor and evaluate) to regulate their learning, and often learn more than other learners who do not engage in the regulation processes (Azevedo et al., 2008; Bannert & Reimann, 2012). To advance research understanding of and facilitate learners' SRL processes, we need to develop novel approaches to measurement and integration of multi-channel data that are used for the study of SRL (Järvelä et al., 2018). This especially to the analysis of micro-level SRL processes, which leads to the investigation of more specific processes within each phase of SRL, e.g., goal setting SRL micro-level process within the planning phase of SRL (Siadat, Gašević & Hatala, 2016). Unobtrusive measures of cognitive, metacognitive, motivational and affective processes can be captured during SRL through log data recorded by digital learning environments (Winne, 2010). However, simple navigational log data or time spent on pages are often not informative enough to study SRL processes (Molenaar & Järvelä, 2014). Hence, we conducted a study that aimed at addressing this problem by enhancing log data with other peripheral data such as mouse movement, mouse click, keyboard stroke, and more interestingly, eye tracking data.

The study used a pre-post design with a 45-minute learning session during which participants (36 university students) were asked to study three topics: 1) artificial Intelligence (the basics of artificial intelligence and how it will influence education in the near future), 2) differentiation in the classroom

(the concept of differentiation explains how teachers can deal with differences between students, and the idea of adaptive learning) and 3)scaffolding learning (as an important way to support students during learning and to adjust to the needs of individual students.). The learning task was to integrate the three topics into a vision essay (300-400 words long) that describes learning in school in 2035. The study used a learning environment (see Fig. 1) with five areas of interest (AOI) zones. The iMotions software system was used to record and synchronize multi-channel data with a unified timeline.

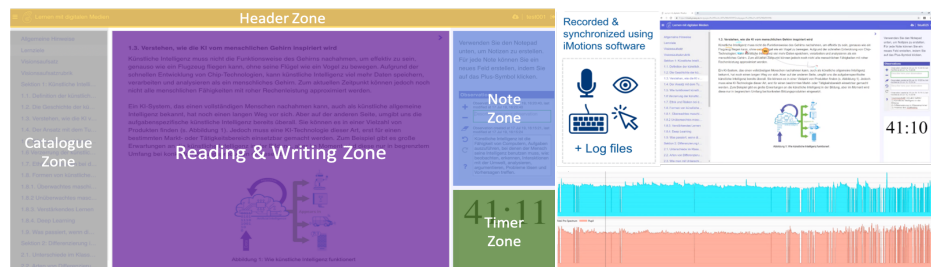


Figure 1: Learning environment (AOI) and iMotions system (synchronizing multi-channel data)

2 MEASUREMENT PROTOCOL FOR MICRO-LEVEL SRL PROCESSES

Based on the framework proposed by Siadat, Gašević & Hatala (2016), we developed a measurement protocol for detection of SRL processes from combined log and eye-tracking data (see Fig. 2). The protocol contains i) rules for identification of SRL processes (e.g., planning) and ii) a log parser which turns raw log data into learning events or alternatively “event-ized” trace data. In order to analyze how eye tracking can provide richer information, as compared to the enhanced log data (here we include the mouse and keyboard events), we built the action library with two separate data channels (log only/log+eye track) (see Table 1). The action library provides the definition of 10 action labels, which are the codes for individual learning actions (e.g., when Learners have a quick glimpse at the timer, we label this action as “TIMER”). The pattern library consists of patterns of sequential actions labelled in the action library (e.g., when Learners have a quick glimpse at the timer during essay writing, we detected the learning pattern as “WRITE_essay to TIMER back to WRITE_essay”), and it was built to map learning patterns with micro-level SRL processes. The pattern library, which included cognition patterns and metacognition patterns, was based on Bannert’s (2007) SRL coding scheme. The detailed pattern library is not shown in this paper due to the length restrictions.

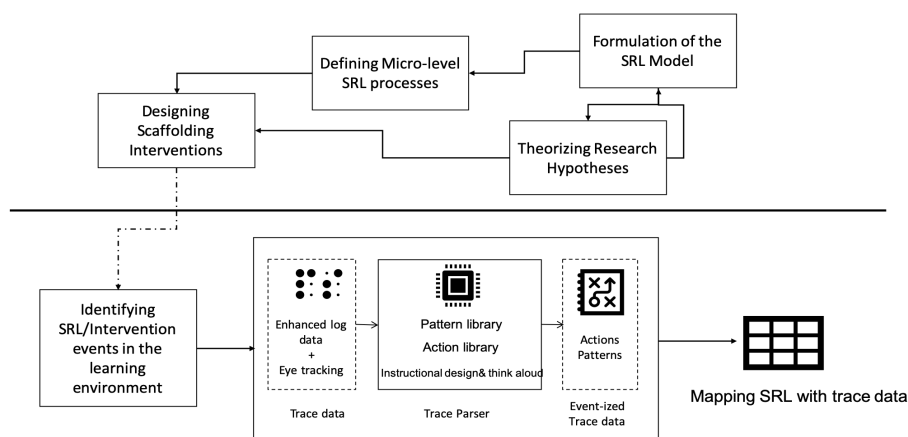


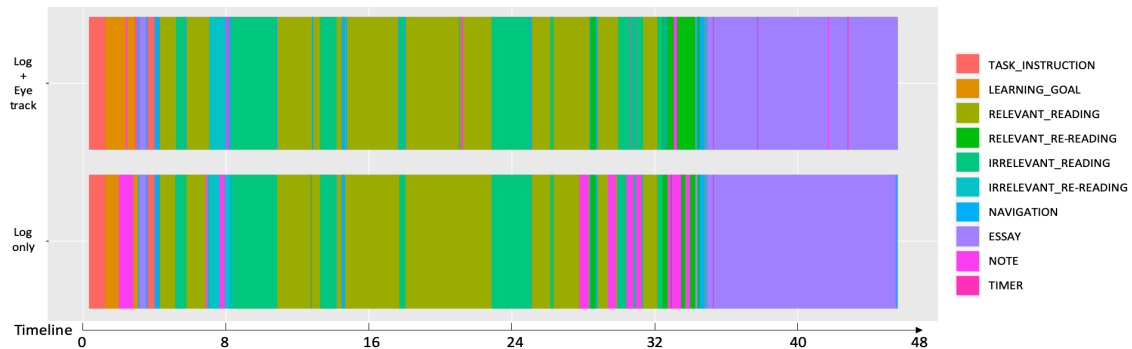
Figure 2: The Measurement Protocol of Integrating Multi-Channel Data

Table 1: Action library for detection of SRL processes from trace and eye-tracking data

Labels	Action definition	Data	Examples
TASK_INSTRUCTION	Learners read or re-read the general instructions page and the essay rubric page	log only	Learners open essay rubric page to understanding the task
		log+eye track	Learners read essay rubric with fixation task requirements
LEARNING_GOAL	Learners read or re-read the learning goals page	log only	Learners open and read learning goal page
		log+eye track	Learners open and read learning goal page with fixation
RELEVANT_READING	Learners read and learn relevant content for the first time	log only	Learners open and read relevant content page (e.g., AI definition)
		log+eye track	Learners read relevant content with fixation in the reading zone
RELEVANT_RE-READING	Learners re-read and review for relevant content which they have read before	log only	Learners re-open “AI definition” page during essay writing
		log+eye track	Learners re-read preceding part of the page with overlap fixation
IRRELEVANT_READING	Learners read pages which are not relevant to the learning goal and essay writing	log only	Learners open and read relevant content page (e.g., Turing Test)
		log+eye track	Learners read irrelevant content with fixation in the reading zone
IRRELEVANT_RE-READING	Learners re-read pages which are not relevant to the learning goal and essay writing	log only	Learners re-open “Turing Test” page after reading other pages
		log+eye track	Learners re-read preceding part of the page with overlap fixation
NAVIGATION	Learners view or glance at catalogue zone or overview page, or quickly navigate through pages	log only	Learners quickly click through pages to overview materials
		log+eye track	Learners fixate at catalog zone after reading through one page
WRITE_ESSAY	Learners write, edit, re-write the essay, or stay in the essay page to think about essay writing	log only	Learners type and write sentences in the writing zone
		log+eye track	Learners fixate at the writing zone without typing
NOTE	Learners add, delete, write, edit or read notes	log only	Learners click in the note zone to create a new note after reading
		log+eye track	Learners fixate at notes they took before during essay writing
TIMER	Learners check timer during the learning task	log only	Learners use mouse click or scroll at the timer zone
		log+eye track	Learners have a quick glimpse at the timer

3 PRELIMINARY RESULTS OF THE CASE STUDY

In order to show some preliminary results, here we use participant P25 as a case study. P25 left 25235 rows of enhanced log data (21,982 mouse moves/clicks/scrolls; 3,019 keystrokes, and 250 BrowserNav/Scrolls), and 7325 rows of fixation data (with more than 1.2 million rows gaze data), in a 45 minutes learning session. All ten labels from the action library (Table 1) were detected based on the enhanced log data of P25: P25 spent approximately 4 minutes in the beginning to read the task instruction and the learning goal, then spent almost 30 minutes to read or re-read the content with note-taking, and finally, spent approximately 10 minutes in the end to write the essay. The timeline of the learning processes is shown in Figure 3, based on “log only” or “log+eye tracking”.

**Figure 3: Learning processes detected from multi-channel dataset**

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

From the “log only” data, we detected micro-level SRL processes: in the planning phase—taking notes while reading the learning goal page (*LEARNING_GOAL* to *NOTE* back to *LEARNING_GOAL*); in the orientation phase—navigating through many reading pages after reading the task instruction page (*TASK_INSTRUCTION* to *NAVIGATION*). We also detected cognition patterns, such as elaboration patterns (e.g., *RELEVANT_READING* to *RELEVANT_RE-READING*) and organization patterns (e.g., *RELEVANT_READING* to *NOTE* back to *RELEVANT_READING*) in the reading stage. However, we were able to find many more detailed SRL processes when adding eye tracking data into enhanced log data, especially more monitoring patterns such as a quick glimpse at the timer to monitoring the time process during writing (*WRITE_essay* to *TIMER* back to *WRITE_essay*).

4 DISCUSSION AND NEXT STEPS

In this study (Project name: FLORA, funded by ORA; BA20144/10-1, NWO 464.18.104, ES/S015701/1), we proposed a measurement approach for detection of theoretically meaningful micro-level SRL processes from enhanced log data and eye tracking data, such as orientation, planning and monitoring. In general, the addition of eye tracking to log data to enrich information about the learning area of interest, greatly improved the measurement of temporal patterns such as checking notes/timer with just a glimpse and without mouse clicks/moves. In order to further triangulate our findings, in the future work we will also integrate think-aloud data into our multi-channel dataset. We will use think-aloud data to shed more light on the measurement of SRL processes, and more importantly, to validate the inferences drawn from the trace data.

REFERENCES

- Azevedo, R., Moos, D.C., Greene, J.A., Winters, F.I., & Cromley, J.G. (2008). Why is externally-facilitated regulated learning more effective than self-regulated learning with hypermedia? *Educational Technology Research and Development*, 56(1), 45-72.
- Bannert, M. (2007). *Metakognition beim Lernen mit Hypermedia*. Münster: Waxmann.
- Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through prompts. *Instructional Science*, 40(1), 193–211. <https://doi.org/10.1007/s11251-011-9167-4>
- Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. A. (2018). What multimodal data can tell us about the students’ regulation of their learning process?. *Learning and Instruction*(in press), <https://doi.org/10.1016/j.learninstruc.2019.04.004>
- Molenaar, I., & Järvelä, S. (2014). Sequential and temporal characteristics of self and socially regulated learning. *Metacognition and Learning*, 9(2), 75–85. <https://doi.org/10.1007/s11409-014-9114-2>
- Winne, P. H. (2010). Improving Measurements of Self-Regulated Learning. *Educational Psychologist*, 45(4), 267–276. <https://doi.org/10.1080/00461520.2010.517150>
- Siadat, M., Gasevic, D., & Hatala, M. (2016). Trace-based micro-analytic measurement of self-regulated learning processes. *Journal of Learning Analytics*, 3(1), 183-214.

Identifying Trigger Regulation Events in Collaborative Learning

Sanna Järvelä

University of Oulu
sanna.jarvela@oulu.fi

Jonna Malmberg

University of Oulu
jonna.malmberg@oulu.fi

Eetu Haataja

University of Oulu
eetu.haataja @oulu.fi

Muhterem Dindar

University of Oulu
muhterem.dindar @oulu.fi

ABSTRACT: In this workshop we will discuss our work on identifying trigger regulation events in collaborative learning. Trigger regulation events are those that allow learners to change the course of their learning by using adaptive strategies. First, we review our empirical evidence about using different data channels to evidence of metacognition, cognition, emotion and motivation in collaborative learning (See Table 1). Second, we discuss how different data channels can reveal challenging learning situations in collaboration, namely ‘trigger regulation events,’ which invite learners for regulation in collaborative learning. Third, we demonstrate a case (seven-weeks multichannel process data collection in high school physics lessons) how combination of psychophysiological (activating situations), contextual data (videos) and learners beliefs (situated self-reports) can provide a theory based lens to capture students’ regulatory actions as well as their regulatory responses to optimize their learning progress (= adaptation).

Keywords: Collaborative learning, self-regulation, socially shared regulation, multimodal data, cognition, metacognition, motivation, emotion

1 INTRODUCTION

Contemporary perspectives view self-regulation as a cyclical complex metacognitive and social process that involves adapting cognition and metacognition, motivation, emotion, and behavior. Regulation is neither a static phenomenon nor a state of the learner, but rather a series of contingencies over time (Zimmerman, 2014; Winne, 2018). In other words, regulation evolves. Earlier approaches to SRL research were successful in identifying students’ general beliefs about their learning as well as generic tactics and strategies they used to regulate learning. What its methodologies could not achieve was to make clear how those actions take place as patterns of

behavior and how learning conditions and patterns of behaviour reciprocally influence each other (Winne, 2014). That is the reason why Learning Analytics (LA) and Adaptive Learning Technologies (ALTs) have failed to achieve their promises.

Theories have been explicit in explaining the role of cognition, motivation, and emotion regulation in SRL (e.g. Schunk & Greene, 2017) and in the empirical verification of each of those components separately during the learning process. While theories emphasize the strong interplay among cognition, motivation and emotion in adaptive or maladaptive learning, empirical work still treats each of them separately because older, limited methodologies are incapable of capturing this the interactive dynamics of learning as a process.

We have been working for theoretical advancement of self-regulation in social learning contexts, namely socially shared regulation in learning (Järvelä, Hadwin, Malmberg & Miller, 2018) in collaborative learning (Järvelä, Malmberg, Haataja, Sobosincki & Kirschner, 2019). So far, we have empirically identified the importance of cognitive (Malmberg, Haataja, Seppänen & Järvelä, 2019), emotional (Järvenoja, Näykki & Törmänen, 2019) and adaptive/maladaptive regulation patterns in collaborative learning (Sobocinski, Järvelä, Malmberg et al. 2020) based on theories of S/SRL (Winne & Hadwin, 1998; Hadwin, Järvelä & Miller 2018). That is, because theories of learning are tools which allow us to track the meaningful events in collaborative learning. However, those events are sometimes invisible and difficult to capture with “naked eye”. That is why we have been implementing multimodal methods to identify when, how and what makes regulation in collaborative learning functional. To do this, we have been relied for our empirical evidence and theoretical understanding of regulation to utilize advanced learning technologies to support learning (Järvelä et al., 2019).

2 AIM

In this workshop we will discuss our work on identifying trigger regulation events in collaborative learning. Trigger regulation events are considered as events that allow learners to change the course of their learning by using adaptive strategies. First, we review our empirical evidence about using different data channels to evidence of metacognition, cognition, emotion and motivation in collaborative learning (See Table 1). Second, we discuss how different data channels can reveal challenging learning situations in collaboration, namely ‘trigger regulation events,’ which invite learners for regulation in collaborative learning. Third, we demonstrate a case how combination of psychophysiological (activating situations), contextual data (videos) and learners beliefs (situated self-reports) can provide a theory based lens to capture students’ regulatory actions as well as their regulatory responses to optimize their learning progress (= adaptation).

Table 1. Summary of the data to be implemented for researching SSRL and evidence obtained in our current research

DATA CHANNEL	THEORETICAL FOCUS			
	COGNITION	METACOGNITION	EMOTIONS	MOTIVATION
LOG FILES	Malmberg et al. (2010)	Pijeira-Díaz et al., (2019)		
EYE TRACKING		Pilot studies		
PHYSIOLOGICAL SENSORS	Sobocinski et al. (2019)	Haataja et al., (2018)	Järvelä et al. (2019)	
360 ° VIDEO+AUDIO	Järvelä et al. (2016)	Näykki et al. (2015)	Järvenoja et al., (2017)	Järvelä et al. (2008)
FACIAL RECOGNITION	Malmberg et al., (2018)		Dindar et al., (2019)	
(SITUATED) SELF-REPORTS*	Dindar et. al (2019)	Haataja et al. (2019)	Järvenoja et al. (2019)	Järvenoja et al, (2012)

(green=strong evidence, orange=accumulating evidence, grey=undirect in combination of other channels,white=no studies/evidence). References have deleted for a review.

3 MULTIMODAL DATA COLLECTION IN HIGH SCHOOL COLLABORATIVE PHYSICS LESSIONS

In this workshop, we elaborate in detail how, why and in what ways we have found evidence for importance of cognition, metacognition, emotions and motivation. The guiding principle in our empirical work is that regulation of learning occurs in authentic learning settings. It can be captured (partly) in laboratory settings, but it often fails to capture the real need for regulated learning. Our recent attempt to capture regulation of learning and how it evolves is collected in secondary school science lessons. The data consists of seven-week multichannel process data collection when high school students' (N = 94) worked collaboratively in groups of three during the physics lessons (See Table 2). Students collaboration was followed with video recordings and through individual level physiological measures. To capture the learning activity in its natural setting and to get multimodal process data related to the different cognitive, emotional and motivational components, the learning session was recorded using four Insta360 Pro video cameras, that were placed in the classroom and separate microphones placed in front of each group.

Video data provides us contextualised data through different channels (voice, facial expressions, interactions) from the different operations shaping the groups' shared as well as group members' individual motivational and emotional states. To capture students' covert physiological reactions during the learning situation, such as students' physiological activation related to emotion and cognition, students' Electrodermal Activity (EDA) and heart rate (HR) were recorded with Shimmer3 GSR + devices. From the EDA measurement, for example, students' general physiological activation level during the learning session as well as short-term emotional responses can be identified (Dawson et al., 2017). In addition, we measured students learning outcomes in both group and individual level. As one of the multiple data sources, we used the 6Q tool implemented in Qridi® to collect students' situation-specific interpretations of their cognition (task understanding and perceived task

difficulty), emotion (valence and activation) and motivation (situational interest) to each collaborative session before and after the collaborative work.

Table 2. Multimodal process data collected in physics collaborative learning tasks

Theoretical focus	Specific construct	Data source	Sample N
Emotions Motivation Cognition Collaboration	e.g. Interaction, self-, co-, and socially shared regulation	Video (Insta360 Pro camera, separate microphone)	7 sessions x 90 min x 30 groups = 212 h
Emotions Metacognition	Physiological arousal & activation Physiological synchrony	Electrodermal activity (Shimmer3 GSR+)	7 sessions x 90 min x 84 students = 583 h
Outcome	Content knowledge	Fact test (Qridi)	7 x 94 = 289/376 responses

4 CONCLUSIONS

Despite we have empirical evidence related on importance of cognition, motivation and emotion in collaborative learning, we have not yet been able to identify how each of these component a) occur and b) intertwine in the context of collaborative learning. So far, our work is (slowly) progressing towards identifying trigger events that set the stage for regulation of cognition, motivation and emotion within collaborative groups and determine the critical signatures of these events that predict success or failure in individual and collaborative learning. To conclude, today, due the advancements of technology there is a promise to capture not only the cognitive activities related on learning, but also the metacognitive, motivational and emotional aspects that determine the quality of collaborative learning. What we aim to achieve is to make clear how regulation of cognition, motivation and emotions takes place as patterns of behavior and how learning conditions and patterns of behaviour reciprocally influence each other (Winne, 2014).

REFERENCES

- Dawson, M. E., Schell, A. M., & Filion, D. L. (2017). The Electrodermal System. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (pp. 217–243). Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781107415782.010>
- Hadwin, A. F., Järvelä, S., & Miller, M. (2018). Self-regulation, co-regulation and shared regulation in collaborative learning environments. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (pp. 83–106). New York: Routledge/Taylor & Francis Group.

- Järvelä, S., Malmberg, J., Haataja, E., Sobosincki, M. & Kirschner, P. (2019). What multimodal data can tell us about the self-regulated learning process? *Learning and Instruction*. <https://doi.org/10.1016/j.learninstruc.2019.04.004>
- Järvenoja, H., Näykki, P. & Törmänen, T. (2019). Emotional regulation in collaborative learning: When do higher education students activate group level regulation in the face of challenges. *Studies in Higher Education*. <https://doi.org/10.1080/03075079.2019.1665318>
- Malmberg, J., Haataja, E., Seppänen, T., & Järvelä, S. (2019). Are we together or not? The temporal interplay of monitoring, physiological arousal and physiological synchrony during a collaborative exam. *International Journal of Computer Supported Collaborative Learning* 14 , 467-490. <https://doi.org/10.1007/s11412-019-09311-4>
- Schunk, D. H., & Greene, J. A. (2017). *Handbook of self-regulation of learning and performance*. Routledge.
- Sobocinski, M., Järvelä, S., Malmberg, J., Dindar, M., Isosalo, A. & Noponen, K. (2020). How does monitoring set the stage for adaptive regulation or maladaptive behavior in collaborative learning? A multimodal analysis. Submitted.
- Winne, P. H. (2014). Issues in researching self-regulated learning as patterns of events. *Metacognition and Learning*, 9(2), 229–237. <https://doi.org/10.1007/s11409-014-9113-3>
- Winne, P. H. (2018). Theorizing and researching levels of processing in self-regulated learning. *British Journal of Educational Psychology*, 88(1), 9–20. <https://doi.org/10.1111/bjep.12173>
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah: Lawrence Erlbaum.
- Zimmerman, B. J. (2013). From Cognitive Modeling to Self-Regulation: A Social Cognitive Career Path. *Educational Psychologist*, 48(3), 135–147. <https://doi.org/10.1080/00461520.2013.794676>



The 10th International

Learning Analytics & Knowledge Conference

Frankfurt, Germany

March 23-27, 2020 #LAK20 

hosted by



Finding Semantic Structure of Content from Gaze Data

Hiroaki Kawashima

University of Hyogo, Kobe, Japan
kawashima@sis.u-hyogo.ac.jp

ABSTRACT: This paper addresses a problem of finding semantic structure of video content using eye-gaze data during video lectures. While video content, such as slide images and transcripts, have useful information to extract semantic relationship between terms or concepts in a course, it is sometimes difficult to evaluate the importance of terms when it appears only a limited number of times. With the support of gaze data, it is expected that important terms looked at frequently and term pairs with important relation compared more than other term pairs are extracted. In this paper, we propose a method to utilize gaze data to generate a concept map by extracting such important term pairs from gaze data and video content information and demonstrate it with a preliminary experiment.

Keywords: Eye-tracking, video lectures, concept maps

1 INTRODUCTION

In this study, we aim at building a framework to track states of each learner in video lectures (e.g., MOOCs) using eye-gaze data and content information. In this framework, we assume that (1) a model can be first constructed by a collection of gaze data of learners captured asynchronously and that (2) each learner's learning states can be estimated online once such a model is available. In this paper, we particularly focus on (1) and addresses a problem of extracting semantic relations from gaze data.

Concept maps of lectures play an important role not only to assess learners' understandings but also to automatically support learners' construction of knowledge by scaffolding them with information that fulfills a knowledge gap. Therefore, automatic extraction of concept maps (as known as concept/knowledge graphs and ontologies) from learning materials, such as semi-structured texts and slides, has been recently attracted interest of Learning Analytics communities (Fillos & Ochoa, 2019; Flanagan, Majumdar, Akcapinar, Wang, & Ogata, 2019). The use of content of textbooks and slides to construct a semantic structure of courses is especially effective when the cost of manual creation is large. This is often the case when concept maps are required not only for the curriculum level but for each of courses, where learning materials are customized together with the content to be learned, and therefore corresponding concept maps need to be refined to describe the content in detail.

While the use of learning materials is a promising approach to automatically generate content-dependent semantic structures, it sometimes depends on the statistics of terms that appear on the content. That is, automatic extraction of concept maps often assumes that important terms or concepts appear frequently in the learning materials. However, this assumption does not always hold when we need a concept map for a specific course or lecture due to the limited amount of learning materials. On the other hand, content-dependent concept maps are expected to be effective to generate feedbacks specific to each lecture.

To overcome this limitation and to apply an automatic construction of concept maps to a variety of learning materials such as online video content, we investigate the use of collection of gaze data obtained from viewers who watched the same lecture video. Each viewer's eye-gaze can be recorded by a gaze tracker as a series of fixations on a screen and converted to a sequence of AOIs (attention-of-interests), which we refer to as *gaze regions*. Therefore, important keywords and pairs of terms with important relations are expected to be looked at frequently by the viewers and are identified even when they appear only once in a slide or are uttered a limited number of times by a lecturer.

The research question of this study is summarized as follows: Does the integrated use of gaze data and video content provide useful information to extract semantic relations? We examine with preliminary results the possibility of using multi-modal analysis for automatic concept-map generation.

2 AUTOMATIC EXTRACTION OF CONCEPT MAPS

Suppose slide data and transcripts are available with video content. We first segment each slide into small regions of words, figures, symbols (e.g., arrows), or mathematical equations. We denote the regions in a slide as $\{R_1, \dots, R_N\}$, where the content of each of the regions is annotated. This step is done manually to focus on the subsequent steps while it is also possible to introduce automatic segmentation of a slide image. The transcripts are processed by a text mining tool called KH Coder (Higuchi, 2016) to extract a list of terms. Connective words and different word forms are checked and fixed manually. Because of the limited amount of text, in this analysis we use word-class filtering together with a threshold of minimum appeared count instead of *tf-idf* to filter-out unimportant terms. While the minimum count is set to two, some manually determined important terms are preserved.

Raw gaze sequences of xy-coordinates are converted to gaze regions $r_1, \dots, r_{k-1}, r_k, \dots, r_K$, where $r_k \in \{R_1, \dots, R_N\}$, and time (or time step) k denotes an ordered number of AOI switches. Meanwhile, we use t_k to describe actual *media time* (physical time whose origin is the start time of a slide) in the video at time step k , where the grid size of t_k is 1 s.

As a simple usage of gaze region sequences, we utilize gaze-region transition probability $P(r_k|r_{k-1})$ and region-time probability $P(r_k|t_k)$. First, both the probabilities are computed, and then regions with high probability (above a given threshold) are extracted. Then, annotated word in the regions is used to compute the co-occurrence of terms. Here, gaze-region transition probability is scan-path statistics in fixations and can be used to extract which term pairs are looked at in order. We take the average of $P(r_k = R_i|r_{k-1} = R_j)$ and $P(r_k = R_j|r_{k-1} = R_i)$ to extract comparison behavior. Meanwhile, region-time probability identifies synchronized fixations. We apply a sliding window method (window size: 10 s, step size: 5 s) to extract term co-occurrence in the temporal intervals.

From the preliminary analysis, we found that the method of mode decomposition (Kawashima, Ueki, & Shimonishi, 2019a/2019b) finds more sparse and meaningful probability distributions. By assuming dynamic change of internal attentional mode ($m_k = 0, 1, 2$) of each viewers, the method extracts mode-dependent transition probability $P(r_k|r_{k-1}, m = 1)$ and region-time probability $P(r_k|t_k, m = 2)$ together with $P(r_k|m = 0)$ (base mode), where mode 1 and 2 can be interpreted as slide-following (region order dependent) and lecturer-following (time dependent) behaviors, respectively. That is, if a learner followed a lecturer's speech and pointers, mode 2 is estimated to be dominant, similar to the concept of "with-me-ness" (Sharma, Jermann, & Dillenbourg, 2014).

3 PRELIMINARY EXPERIMENTS

In this pilot study, we use gaze data obtained from 11 participants who watched “Statistics 2” from JMOOC gacco (<https://gacco.org/>). Figure 1 shows a ground-truth concept map (directions are omitted) determined manually by watching the video consisting of three-page slides. Node terms were selected from the term list used in the text analysis. Two ground-truth concept maps were respectively prepared by two lecturers who are teaching “Statistics” to University students. One of the two lecturers (denoted by “lecturer 2”) is the author, and therefore its related results are shown for reference in what follows.

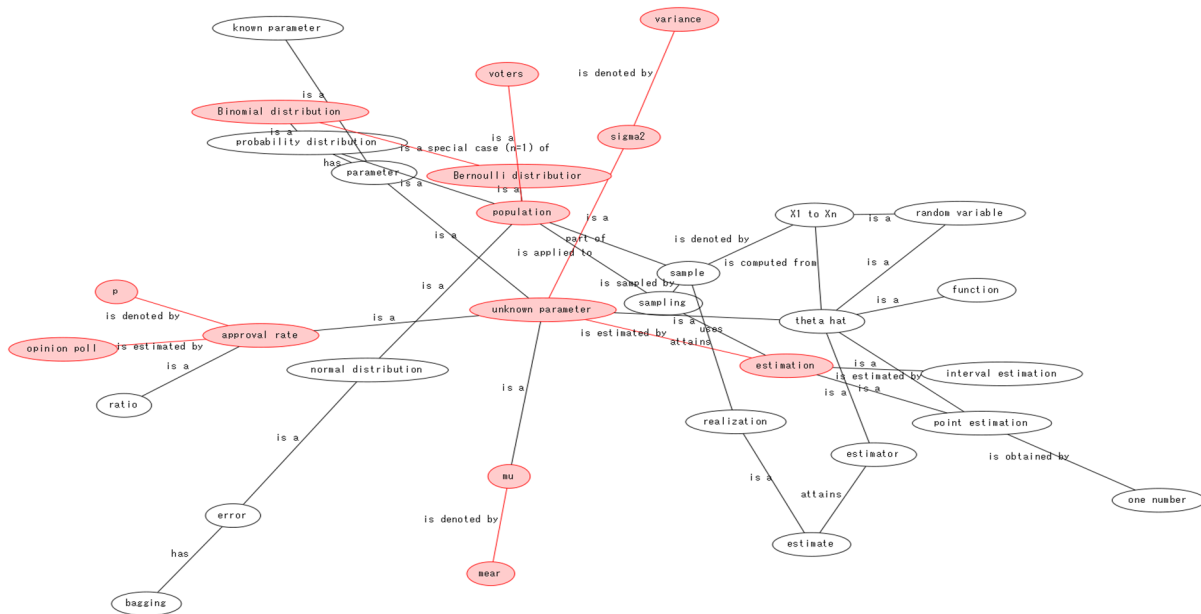


Figure 1: A ground-truth concept map for the three-slide video content created by lecturer 1 and an example of detected relations from gaze data during watching of one of the slides (colored)

For the evaluation, we used the second slide, obtained the term networks (weighted undirected graphs) using each of the following methods, and compared them with the ground-truth concept maps: (1) **Co-TS** (word co-occurrence in transcripts), (2) **R-Prox** (region proximity in a slide), (3) **Mode1** (gaze-region transition), (4) **Mode2** (gaze synchronization), where (1) and (2) are baseline methods. For (3) and (4), we also introduced term filtering using the term list found through the transcript mining, denoted respectively as **F-Mode1** and **F-Mode2**. Jaccard similarity was used to compute co-occurrence of terms for **Co-TS**, **Mode2**, and **F-Mode2**. For **R-Prox** and **Mode1** (**F-Mode1**), proximity (distance) of regions and probabilities are respectively used for edge weights.

Figure 2 shows the accuracy of the predicted semantic structures (term networks), where the values are the ratio that top-30 edges ordered by the weights are also included in each of the two ground-truth concept maps (horizontal axis: not only direct neighbor but larger path length between two nodes are also allowed). In this experiment, we do not evaluate the coverage of the ground truth. Although the absolute values of the accuracy are affected by how the ground truth is prepared, we can see that the gaze-oriented methods (Mode1 and Mode2) are higher than the baselines (Co-TS and R-Prox). Moreover, the term filtering further improves the accuracy as shown in the figure, and this is consistent between the evaluation on the two ground-truth concept maps.

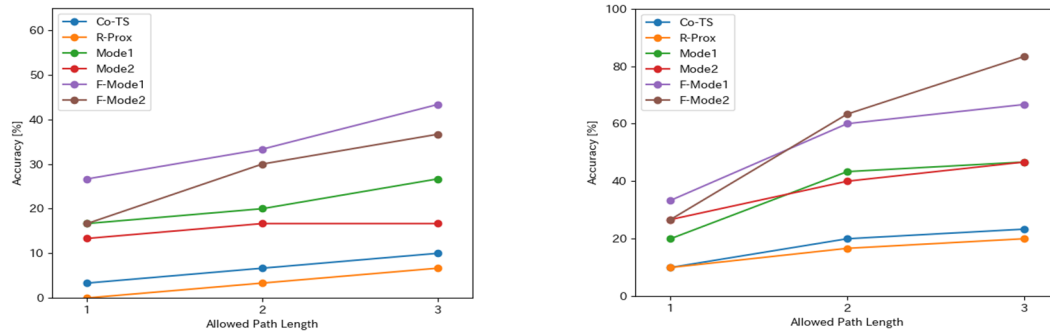


Figure 2: Accuracy of term-network extraction. The values are the accuracies of edge prediction of the ground-truth concept map annotated by lecturer 1 (left) and lecturer 2 (right)

4 CONCLUSIONS

This result suggests that eye-gaze data have some important clues to further improve the result of the construction of the concept maps while we still need to investigate more sophisticated automatic extraction methods of concept maps from content information itself (e.g., Fillos, 2019) and several manual steps need to be replaced. As for data collection, we are now planning to collect larger dataset for the evaluation. We consider that the assumption of gaze measurement is not unrealistic since eye tracking devices and techniques including camera-based method (Zhang, Sugano, Fritz, & Bulling, 2015) are now improving. While we focus on the first problem: “how to utilize eye-gaze data for concept maps construction,” in future we will address the second problem: “how to estimate learners’ states and styles” by extending the present method and applying other model-based method (e.g., Kawashima, 2019b) to trace an individual learner’s cognitive states from multimodal data.

Acknowledgements: This work was supported by JSPS KAKENHI Grant Number JP19H04226, JST PRESTO Grant Number JPMJPR14D1, and the Telecommunications Advancement Foundation.

REFERENCES

- Fillos, A. and Ochoa, X. (2019). Semi-Automatic Generation of Intelligent Curricula to Facilitate Learning Analytics. *The 9th International Conference on Learning Analytics & Knowledge (LAK)*, 46–50.
- Flanagan, B., Majumdar, R., Akcapinar, G., Wang, J., & Ogata, H. (2019). Knowledge Map Creation for Modeling Learning Behaviors in Digital Learning Environments. *Predicting Performance Based on the Analysis of Reading Behavior, the Companion Proceedings of LAK19*. 422-430.
- Higuchi, K. (2016). A Two-Step Approach to Qualitative Content Analysis: KH Coder Tutorial Using Anne of Green Gables (Part I). *Ritsumeikan Social Science Review* 52(3): 77–91.
- Kawashima, H., Ueki, K., & Shimonishi, K. (2019a). Modeling Video Viewing Styles with Probabilistic Mode Switching. *3rd Multimodal Learning Analytics Across Spaces (CrossMMLA) Workshop, the Companion Proceedings of LAK19*, 526–533.
- Kawashima, H., Ueki, K., & Shimonishi, K. (2019b). Model-Based Analysis of Gaze Data During Video Lectures. *27th International Conference on Computers in Education (ICCE 2019)*, 81–86.
- Sharma, K., Jermann, P., & Dillenbourg, P. (2014). “With-me-ness”: A Gaze-Measure for Students’ Attention in MOOCs. *International Conference of the Learning Sciences (ICLS)*, 1017–1022.
- Zhang, X., Sugano, Y., Fritz, M., & Bulling, A. (2015). Appearance-based gaze estimation in the wild. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 4511–4520.

Mobile Multimodal Learning Analytics Method to Foster Student Self-Regulated Learning

Mohammad Khalil

Centre for the Science of Learning & Technology (SLATE), University of Bergen
mohammad.khalil@uib.no

Olga Viberg

KTH Royal Institute of Technology
oviberg@kth.se

ABSTRACT: The aim of this workshop paper is to propose Mobile Multimodal Learning Analytics Methodology (MOLAM). The methodology is suggested to be developed through the lenses of multidisciplinary and multichannel data research approaches, based on the theoretical foundations of Self-Regulated Learning (SRL). MOLAM is theory supported, driven by learning analytics, learner-centered focused, and mobile technology utilized. We argue that MOLAM will have a potential to support learners, teachers and researchers in their understanding and their further fostering of student SRL in formal and informal learning environments.

Keywords: Learning analytics; mobile multimodal learning analytics methodology; multichannel data; self-regulated learning

1 BACKGROUND AND METHOD PROPOSAL

The aim of this workshop paper is to discuss our proposal of a research area which we call Mobile Multimodal Learning Analytics Methodology (MOLAM) to trace, interpret and support students' development of self-regulated learning (SRL) strategies, skills and knowledge. While the focus of the learning analytics research varies (Khalil & Ebner, 2016; Viberg et al., 2018), increasing research attempts have recently targeted the area of self-regulated learning (SRL; Viberg, Khalil & Baars, 2020).

SRL refers to how learners steer their own learning (Wong et al., 2019). It is a broad process-oriented concept that encompasses motivational, metacognitive, cognitive, affective, and behavioural aspects of learning (Panadero, 2017). It is a well researched area. Yet, related research attempts have largely focused on understanding SRL activities as static learning processes by using subjective self-reported assessment measures such as surveys, self-reported or think-aloud methods. However, SRL is not only a static process; it increasingly evolves (Sedryakyan et al., 2018), and even though considerable theoretical and conceptual progress has been made with respect to regulation in learning, there has been "little progress in developing methods to make the primary invisible mental regulation processes [...] visible and thus measurable and ultimately interpretable" (Noroozi et al., 2019, p. 299). This is critical since earlier research has shown that many students possess poor SRL practices, including the ability to accurately calibrate their own learning processes (Dunlosky & Rawson, 2007). Further, it has been found that without instructional support, students

may overestimate their understanding of learning materials (Baars et al., 2018; Thiede et al., 2009). All this suggests that to be able to successfully support learners in their development of SRL, we need to better understand the evolving continuous nature of the multifaceted learning processes that constitute their SRL. To fill this gap, we propose the innovative MOLAM that is argued to benefit from the use of mobile learning analytics (Aljohani & Davis, 2012) in a combination with multimodal data collection, the analysis of which has earlier been found beneficial for the understanding of students' SRL processes (Järvelä et al., 2019).

MOLAM will be developed based on the theoretical grounding of SRL (e.g., Zimmerman, 1990; Zimmerman & Schunk, 2011; Panadero, 2017). MOLAM brings potential in terms of generating multichannel mutually constituting process-oriented SRL data. Firstly, the use of mobile technologies in combination with learning analytics (LA) is an under-researched area within the field of LA and educational data mining (Shorfuzzaman et al., 2019). The area is challenging because of the characteristics offered by mobile devices. For example, there is a large amount of temporal process-oriented learner data that can be collected with a different flavour than those existing in web-based systems. Mobile learning provides possibilities of having localized data and information collected from numerous learning sessions (Tabuenca et al., 2015) continuously occurring across formal and informal educational settings. Combined with learning analytics, or Mobile Learning Analytics (Aljohani & Davis, 2012), it is understood as "the collection, analysis and reporting of the data of mobile learners, which can be collected from the mobile interactions between learners, mobile devices and available learning materials; it is also supported by the pre-registered data about learners in different university systems" (p. 71).

Secondly, the use of multimodal data collection together with LA is getting increased attention during the last years (e.g., Dindar et al., 2019; Järvelä et al., 2019). In mobile technologies, data, as a multimodal data source, can be generated through built-in services and sensors such as GPS-location, wifi signals, speech input, and fingerprints. Within educational context, mobile multimodal data will offer new insights on state-of-the-art approaches for learning behaviours including metacognitive and cognitive aspects of learning which by then imply SRL processes.

Based on the fact that SRL cannot only be learnt, but also taught (Raaijmakers et al., 2018), the proposed method will aid three groups of stakeholders, namely students, teachers, and researchers. To support students in their development of SRL, we argue that they will benefit from, but not limited to, two support tools. On the one hand, they will benefit from the use of specially developed or adapted software/apps - that are easily accessible through their mobile devices (e.g., smartphones and/tablets) - aiming at explaining and practicing SRL in selected learning settings. Through the use of their own mobile technologies, in which we propose to integrate mechanisms for collecting multimodal data, a relevant process-oriented multichannel data will be collected. On the other hand, by applying mixed-methods mobile multimodal learning analytics approaches, for instance including sensors data together with fingerpress stream, the results will be used to develop a student-facing learning dashboard - a digital tool that visualises students' SRL processes, based on a multichannel data stream (including student log activity data from the adapted SRL software use and multimodal data), with the overall goal to facilitate the development of students' self-regulation. Making SRL processes continuously visible for learners will improve their ability to self-regulate their learning.

The findings of the mobile multimodal learning analytics will also be employed to aid teachers through the development of a teacher-faced learning dashboard that will visualise students' SRL processes, both on individual- and group level. This dashboard will assist teachers not only in their understanding of students' SRL processes but also in designing and practicing relevant teaching activities aiming at further fostering students' SRL in educational settings and providing adequate support.

Finally, to aid researchers to trace and interpret students' SRL activities through a process-oriented approach, we suggest that a graphical user interface that will facilitate data visualisation and processing, thus offering new opportunities for researchers to travel through the learner data and its characteristics needs to be developed. This will contribute to a deeper understanding of the underexplored role of self-regulation in the mobile learning research field (Viberg & Andersson, 2019) and a further theoretical development of the SRL research area.

2 FUTURE RESEARCH DIRECTION

At the triangle connection between learning analytics, mobile technology, and SRL could offer new methods that are not primarily based on frequently used subjective assessment measures, such as learners' perceptions and attitudes towards their SRL, but on the actual use of SRL strategies and related SRL activities during learning. Future directions should be initiated by developing mobile applications that aim at fostering student SRL in increasingly emerging online learning settings and that use multichannel data to measure and track ongoing SRL activities. MOLAM can then be utilized to provide adaptive SRL interventions that will aid teachers to support their learners. We believe unfolding a structure and a framework for MOLAM will be promising in the LA research area considering both the ethical and privacy aspects of students learning behaviour.

REFERENCES

- Aljohani, N. R., & Davis, H. C. (2012). Significance of learning analytics in enhancing the mobile and pervasive learning environments. In *2012 Sixth International Conference on Next Generation Mobile Applications, Services and Technologies* (pp. 70-74). IEEE.
- Baars, M., Leopold, C., & Paas, F. 2018. Self-explaining steps in problem-solving tasks to improve self-regulation in secondary education. *Journal of Educational Psychology*, 110, 578–595. DOI: 10.1037/edu0000223
- Dindar, M., Malmberg, J., Järvelä, S., Hataaja, E., & Kirschner, P. (2019). Matching self-report with electrodermal activity data: Investigating temporal changes in self-regulated learning. *Education and Information Technologies*. DOI: 10.1007/s10639-019-10059-5
- Dunlosky, J., & Lipko, A. R. 2007. Metacomprehension a brief history and how to improve its accuracy. *Current Directions in Psychological Science*, 16, 228-232.
- Järvelä, S., Malmberg, J., Haataja, E., Sobosincki, M. & Kirschner, P. (2019, in press). What multimodal data can tell us about the self-regulated learning process? *Learning and Instruction*. DOI:[10.1016/j.learninstruc.2019.04.004](https://doi.org/10.1016/j.learninstruc.2019.04.004)

- Khalil, M. & Ebner, M. (2016). What is Learning Analytics about? A Survey of Different Methods Used in 2013-2015. In *Proceedings of Smart Learning Conference*, Dubai, UAE, 7-9 March, 2016 (pp. 294-304). Dubai: HBMSU Publishing House.
- Noroozi, O., Alikhani, I., Järvelä, S., Kirschner, P., Seppänen, T., & Juuso, I. (2019). Multimodal data to design visual learning analytics for understanding regulation of learning. *Computers in Human Behavior*. DOI: 10.1016/j.chb.2018.12.019
- Panadero, E. (2017). A Review of Self-regulated Learning: Six Models and Four Directions for Research. *Frontiers in Psychology*, 8, 422. DOI: 10.3389/fpsyg.2017.00422
- Raaijmakers, S. F., Baars, M., Schaap, L., Paas, F., Van Merriënboer, J., & Van Gog, T. 2018. Training self-regulated learning skills with video modeling examples: Do task-selection skills transfer?. *Instructional Science*, 46(2), 273-290. DOI: 10.1007/s11251-017-9434-0
- Shorfuzzaman, M., Hossain, M. S., Nazir, A., Muhammad, G., & Alamri, A. (2019). Harnessing the power of big data analytics in the cloud to support learning analytics in mobile learning environment. *Computers in Human Behavior*, 92, 578-588.
- Tabuenca, B., Kalz, M., Drachsler, H., & Specht, M. (2015). Time will tell: The role of mobile learning analytics in self-regulated learning. *Computers & Education*, 89, 53-74.
- Thiede, K. W., Griffin, T. D., Wiley, J., & Redford, J. S. 2009. Metacognitive monitoring during and after reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 85-106). New York: Routledge.
- Viberg, O., & Andersson, A. (2019). The role of self-regulation and structuration in mobile learning. *International Journal of Mobile and Blended Learning*. DOI: 10.4018/IJMBL.2019100104
- Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behaviour*. DOI:10.1016/j.chb.2018.07.027
- Viberg, O., Khalil, M., & Baars, M. (2020). Self-regulated learning and learning analytics in online learning environments: A review of empirical research. In proceedings of LAK'20. Frankfurt, Germany.
- Wong, J., Khalil, M., Baars, M., de Koning, B. B., & Paas, F. (2019). Exploring sequences of learner activities in relation to self-regulated learning in a massive open online course. *Computers & Education*. DOI: 10.1016/j.compedu.2019.103595

Fifth Grade Students' Problem-Solving Strategies and "Aha! Moments" in Authentic Informal STEM Environment

Yaoran Li^{1*}, Vitaliy Popov¹², Perla Myers¹, Joi Spencer¹, Odesma Dalrymple¹, and
Scott Lundergan¹

¹University of San Diego

²University of Michigan

*yaoranli@sandiego.edu

ABSTRACT: Both cognitive and learning scientists have made efforts to unravel the "secret" of "Aha! moments". This exploratory study aims to use multimodal data to identify Aha! moments in a paper folding activity and a math pencil puzzle activity during a summer enrichment program for fifth-grade students. Video and audio data of 12 students engaged in the paper folding activities and 8 students engaged in the math pencil puzzle tasks were analyzed using ChronoViz (i.e., a multimodal data analysis software). This exploratory study provides rich insights for future design of key features to be included in the development of auto-detectors for Aha! moments and problem-solving strategies. During the multi-channel LAK'20 workshop, we will share our finding of the relations between "Aha! moments" and students' problem-solving strategies. Furthermore, we will discuss the opportunities and challenges in scaling and automating the data collection, annotation, and analysis of multimodal data in authentic learning settings as well as potential measurement biases that need to be taken into considerations in the future research and development efforts.

Keywords: Aha! moment, problem-solving strategy, multimodal data, informal STEM environment

1 INTRODUCTION

"Aha!", or Eureka, moment happens when a person during a problem-solving process suddenly realizes an important insight or a seemingly simple solution (Jiout & Newcombe, 2015; Kounios & Beeman, 2015). Eureka is important because it is a clear indication of learning gain and some of the major creations or scientific breakthroughs happened during the eureka moments (Kounios & Beeman, 2015). A deeper understanding when and how Aha! moment happens in natural learning settings also has important educational implications.

Both cognitive and learning scientists have made efforts in characterizing Aha! moments and understanding when and how Aha! moments happen in the past two decades. Researchers on Aha! moments in the cognitive science community commonly refer to Aha! moments as the experience with sudden insights and have been focused on the cognitive aspects of and the neuroscience connections underlying this type of experience in the context of creative problem solving. In these studies, researchers deployed certain complex problems such as compound remote associate (CRA) problems and asked participants to self-rate their problem-solving process as more "strategy" or more "insights"; and the awareness of how the results were derived or a deliberate decision process were considered to be more "insights" (e.g., Bowden & Jung-Beeman, 2003; Zedelius & Schooler, 2015). Thus the "insight" only measure people's process of a single-step problem; so, it is unclear how people experience "Aha! moments" when facing an authentic problem when multiple steps are necessary. Learning scientists on the other hand, paid more attention on the affective components

as well as the demonstration in actions of Aha! moments. For example, Craig et al., (2004) proposed that when students have experienced negative affect (e.g., frustration), but still actively search for new insights, they would experience the Aha! moment when the insights are profound (Craig, Graesser, Sullins, & Gholson, 2004). They coded Aha! moment when “participants were observed to transfer from a state of confusion to a state of intense interest, as manifested by typing in answers very quickly after a period of inactivity.” Students’ affective state was manually recoded by a 30 second observation every 5 minutes; but Aha! moment was rarely observed in this study. However, using an emote-aloud protocol, D’Mello et al., (2006) reported a higher frequency of Aha! moments. In their study, Eureka was defined as “a feeling used to express triumph on a discovery” and Eureka was identified as the fourth most likely happened affect among the eight affects provided.

Researchers have been focusing on the different aspects of the Aha! moments and the definitions of Aha! moments varied across studies. Given that it is a complex yet valuable learning status, an integrated approach can be beneficial to reveal the full characteristics of Aha! moments and help differentiate Aha! moments from the strategy selection and moments of purely sudden realization or delight. Also, by separating “Aha! moment” with the problem-solving strategy and investigate this concept in different authentic problem-solving processes, we will be able to better understand how different strategies are related to “Aha! moments”. Also, most of the existing studies used college or high school-age students, so there is lack of understanding of how young students experience Aha! moments.

2 METHODS

2.1 Data Collection and Equipment

This current study serves as an exploratory study to understand the nature of aha moments in authentic K12 informal STEM learning context, with video, audio, and survey data collected during an NSF-funded STEM summer program with fifth-grade students. Data of 12 students engaged in the paper folding activities and 8 students engaged in the math pencil puzzle tasks are included this study (Fig. 1).

The study and equipment setup is shown in Figure 1. Two Zoom Q8 Handy Video Recorders were placed on the short side of each table; one was set about 45 degrees to capture the hand movement, and students’ work, and the other was set horizontal to the table to better capture students’ facial expression. Two Pressure Zone Microphones (PZMs) were set on each table. One mic was facing the left side of the table and the other was facing the right side of the table. Because they are the directional mics, this set up helped to differentiate students’ voices. Each camera also has a mic that capture all speeches of the table. We found that the quality of the audio data captured by these mics were much worse quality than the PZMs, so those data were not used during the analysis.

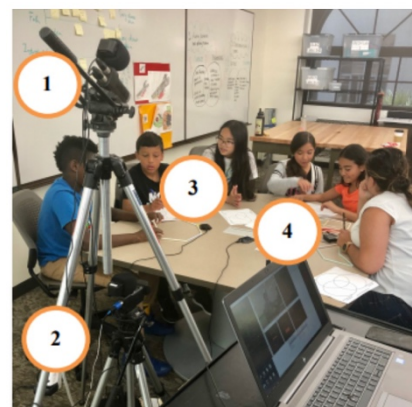


Figure 1. Setup of the multimodal data collection
1: top video recorder, 2: bot video recorder,
3: left PZM, 4: right PZM

2.2 Data Analysis

By integrating the cognitive and affective aspects of “Aha! moment” proposed existing literature, we propose that an Aha! moment should satisfy two main conditions:

- 1) New insights: the sudden realization confers a new/novel perspective or solution to a problem; indicated by actions related with the new insight;
- 2) Dynamic affective transition: the sudden realization was accompanied with an expression of emotion changes from surprise/confusion to delight; indicated by changes in facial expression, body posture, and or spontaneous speech such as “oh” “ah”, “yeah”.

For students’ strategy selection, we code the multimodal data by using the following coding scheme:

- Self-driven: the student did not ask for help and the sudden insight was not depended on peers’ or teachers’ hints or support.
- Ask for help (peer): the student asked for teacher’s help or the teacher provided hints or support during the problem-solving step.
- Ask for help (teacher): the student asked for teacher’s help or the teacher provided hints or support during the problem-solving step.
- Trial and error: the student tried to solve the problem by randomly trying out different potential solution, which indicated by rapid trials.
- Strategic: the students tried to solve the problem by retrieving prior knowledge or by observing the link to prior steps or other information, which indicated by a prolonged observation or thinking process before making a trial.

Because there was minimum interaction between tables during the folding activity, we conducted the data analysis by table. We firstly synchronized the multiple video and audio data files and inputted the time-coded data of each table into ChronoViz (<http://chronoviz.com/>), a software for multimodal analysis. This software is particularly useful for the analysis of video with other types of data sources. The research team conducted manual coding on the multimodal data aiming to gain some insights for the feature engineering of an auto-detector development planned for the future.

3 PRELIMINARY RESULTS

The research team has completed the analysis of the data of the paper folding session and will complete the analysis of the math pencil puzzle task before the conference. From the paper folding task data, we found that there were 32 instances of sudden realization moments detected across the 12 students. There is a large individual difference in the frequency of the detected sudden realization moments (min=0; max=11) with more sudden realization moments were identified for boys than for girls. In our sample, boys were more expressive in both emotion and language than girls.

During the annotation of the multimodal data process, we found that the sudden realization moments could be easily detected through students’ spontaneous speech such as “oh”, “yeah”, “ah” in a combination with their facial expression of wide open eyes, a dropped jaw, or body posture of raising hands/arms. However, when determining whether the sudden realization is an Aha! moment,

more contextual information was taken into considerations, based on our hypothesized conditions (i.e., new insight, assimilation of the insight, and dynamic affective transition).

We found that for the Aha! moments that we identified; the learners all had some sort of mental preparation of a problem. This may be demonstrated as struggling with the task at hand or increased interest toward finding a solution, which is aligned with the research on self-regulated learning (Greene & Azevedo, 2010; Zimmerman, 1990). Self-regulated learners typically prepare themselves by activate existing knowledge and set their learning goals for the task at hand. It would be interesting to investigate how students reflect and solidify/ assimilate the newly formed insight to optimize the process in the future (Molenaar & Järvelä, 2014).

One notable key-differentiating element of ah-ha moments from oh/oh-yeah moments is role of agency involved in the learning process. If there is a teacher-led process, such as during the acquiring new skill phase, the Aha! moments may not occur. This was evident in the current study by three out of five Aha! moments detected occurred during the phase 3 when students were prompted to assemble a cube, and no Aha! moments were detected in phase 1 when students were instructed to fold their first module. Also in the two Aha! moments detected in phase 2, they all happened when students tried to use heuristic approach to find the right fold, instead of trying to directly retrieve a forgotten step. When students used direct retrieval through gestures, peeking others, or asking teachers, a sudden realization sometimes “oh/oh-yeah” moments happened but lack of an expression of delight. When it is a routine task, such as repeated practice on the mastered folds, neither “Aha!” or “oh/oh-yeah” moments occurred.

In the two Aha! moments detected in phase 2, they all happened when students tried to use a more heuristic approach to find the right fold, instead of trying to directly retrieve a forgotten step which might seem to be more efficient. When students used direct retrieval through gestures, peeking others’ model, or asking teachers, a sudden realization may happen but many times lack of an expression of delight, so they are classified as the Oh moments. Also, when it is a routine task, such as repeated practice on the mastered folds, neither “Aha!” or “oh/oh- yeah” moments occurred. These findings are in aligned with Moore et al., (2015) finding that low probability of knowing before answering and high probability of guessing predicted the spikiness of the learning probability.

Although with a small sample size, this exploration study of the nature of Aha! moments in authentic K12 informal learning environment using multimodal analysis provides rich insights for the future design of key features to be considered in machine learning model of Aha! moments as well as in the development of sensor-free auto-detectors. The findings imply that Aha! moments may be effectively extrapolated through automated analysis of spontaneous speech, affective dynamics, as well as the actions.

In the workshop presentation, we also plan to show how Aha! moments are manifested in multi- channel data, the antecedents of Aha! moments, the potentials to automate the annotation process using Natural Language processing and computer vision, and the implications to the use of multichannel data and learning analytics in K8 STEM education.

REFERENCES

- Bowden, E. M., & Jung-Beeman, M. (2003). Aha! Insight experience correlates with solution activation in the right hemisphere. *Psychonomic bulletin & review*, 10(3), 730-737.
- Craig, S. D., Graesser, A. C., Sullins, J., & Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning. *Journal of Educational Media*, 29, 241-250.
- Greene, J. A., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational psychologist*, 45(4), 203-209.
- D'Mello, S. K., Craig, S. D., Sullins, J., & Graesser, A. C. (2006). Predicting affective states expressed through an emote-aloud procedure from AutoTutor's mixed-initiative dialogue. *International Journal of Artificial Intelligence in Education*, 16(1), 3-28.
- Jirout, J. J., & Newcombe, N. S. (2015). Building blocks for developing spatial skills: Evidence from a large, representative US sample. *Psychological science*, 26(3), 302-310.
- Kounios, J. & Beeman, M. (2015). *The Eureka Factor: Aha moments, creative insight, and the brain*. New York: Random House Publishing.
- Kounios, J., Frymiare, J. L., Bowden, E. M., Fleck, J. I., Subramaniam, K., Parrish, T. B., & Jung-Beeman, M. (2006). The prepared mind: Neural activity prior to problem presentation predicts subsequent solution by sudden insight. *Psychological science*, 17(10), 882-890.
- Lindström, P., & Gulz, A. (2008). Catching Eureka on the Fly. In AAAI Spring Symposium: *Emotion, Personality, and Social Behavior* (pp. 65-71).
- Molenaar, I., & Järvelä, S. (2014). Sequential and temporal characteristics of self and socially regulated learning. *Metacognition and Learning*, 9(2), 75-85.
- Moore, G. R., Baker, R. S., & Gowda, S. M. (2015). The Antecedents of Moments of Learning. In *CogSci*.
- Zedelius, C. M., & Schooler, J. W. (2015). Mind wandering "Ahas" versus mindful reasoning: alternative routes to creative solutions. *Frontiers in Psychology*, 6, 834.
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational psychologist*, 25(1), 3-17.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant # AISL-1713547.

Challenges in Multichannel Data Discovery and Integration for Monitoring Performance in Self-Regulated Learning

Shashi Kant Shankar

School of Digital Technologies

Tallinn University, Tallinn

shashik@tlu.ee

ABSTRACT: In the context of Self-Regulated Learning (SRL), monitoring of learners' performance by processing learning evidence from multiple channel is sought quite often from educational practitioners and researchers. However, the data intensiveness due to multiple channels brings severe data processing challenges. We propose to use a Data Value Chain in the second phase of SRL (i.e. use strategies & monitor performance) as a conceptual tool to breakdown the data intensiveness and systematically model the data processing activities. In this workshop paper, we highlight challenges associated with data discovery and integration activities. Finally, we plan to discuss such challenges and possible approaches towards envisioned solutions from the workshop participants.

Keywords: Self-Regulated Learning, Data Discovery, Data Integration, Multichannel Data, Data Value Chain

1 SELF-REGULATED LEARNING & MULTICHANNEL DATA

The importance of personal initiative in learning attracts educational researchers towards Self-Regulated Learning (SRL) (Zimmerman, 2005). Learners are metacognitively, motivationally, and behaviorally active participants in their learning. These three conceptualizations of SRL learners are modelled as a three-phase cyclic process (Zimmerman, 1990). First, learners plan and set their own learning goals. To meet their goals, they use different learning strategies by their own, which may span through diverse learning theories, spaces, resources, materials, and environments. They monitor their learning performance – second phase of SRL. Finally, they reflect on their monitored performance and adapt the learnt lessons in their next iteration of planning and setting up the learning goals.

In most of the SRL situations, learners end up using multiple channels (e.g., various digital learning tools, platforms like Learning Management Systems (LMSs), and video recordings) to meet their self-esteemed goals (Dabbagh, 2012). Recent technological advancements enable to streamline those multiple channels to capture learning evidence. The integration and analysis of captured evidence from multiple channels can help SRL learners for monitoring their performance by following a data-driven approach. Moreover, the analytical results can be presented in form of visualized reports to support the SRL learners for reflection and adaption by following the principle of evidence-based decision making.

Recent research works (Blikstein, 2013; Worsley, 2012) suggest to integrate multimodal evidence of learning under the umbrella of Multimodal Learning Analytics (MMLA). They highlight the

complexities at different levels in the journey to process raw multimodal data for finding meaningful patterns for multiple educational stakeholders. For example, easing up the setup process of the tools in the learning context so that quality evidence can be tracked which can further lead the educational practitioners for easy adoption of data-driven practices without facing the technical complexity. Moreover, there are multiple challenges in the journey of data exploitation. For example, there is an involvement of different data processing activities in the journey. Moreover, each of the data processing activities involves different data processing steps. This makes multimodal and multichannel projects data-intensive in order to enhance and optimize educational practices. In addition to the data-intensiveness, there are other challenges in the process to integrate multimodal data like involvement of multiple educational stakeholders, and the veracity of educational data like cognitive practices, pedagogical decisions, and contextual information of the learning scenario.

A recent review uses a Data Value Chain (DVC) (Shankar, 2018) as a conceptual lens to analyze the existing multimodal systems to systematically model the involved data processing activities. The used DVC includes seven data processing activities (collect & annotate, prepare, organize, integration, analysis, visualization, and decision-making) under three groups (data discovery, integration, and exploitation) in form of a value chain. This DVC has its root in Big Data and Data Mining where the seven involved activities need to be carried out in order to mine raw data for finding out meaningful patterns and uncover hidden insights. The review results highlight that there are uneven support to the seven data processing activities of the DVC by the nine reviewed systems. Moreover, the least supported data processing activities are data preparation, organization, visualization and decision-making. Moreover, another research work (Shankar, 2019) reveals the need to specialize the existing DVC for educational data processing where the input is 1) educational data, and 2) multichannel data. This work uses the existing DVC as a conceptual tool to model the involved data processing activities of four authentic learning scenarios. Further, the requirements of the envisioned DVC (i.e., specialized DVC for the processing of educational data) are extracted by interviewing the stakeholders involved in those four scenarios.

In this workshop paper, we follow the similar approaches as mentioned in the above paragraph and position the use of DVC in the second phase of the SRL i.e., ‘use strategies and monitor performance’ (see Section 2). Section 3 highlights major challenges associated with data discovery and integration activities in SRL context. We close the paper with discussion points in Section 4.

2 USE OF DVC IN THE SECOND-PHASE OF SRL

We propose to use the DVC as a conceptual tool in order to breakdown the data-intensiveness and model the data processing activities. The Figure 1 illustrates the three-phase cyclic process of SRL, and the use of DVC in the second phase. The tools generate multiple channels which are streamlined under the umbrella of data discovery activities (collect and annotate, prepare, and organize). This turns into heterogeneous datasets, which are labelled, prepared, and organized. The data integration step aims to integrate the heterogeneous datasets into one dataset. This integrated dataset represent the coherent view of multichannel data in SRL. Finally, the integrated dataset can be processed under data exploitation umbrella of DVC for presenting the visualized results to the SRL learners. SRL learners use those visualizations to monitor their performance and they can reflect

by the sense making of the visualized reports. Finally, they can adapt the reflections in their next iteration of the SRL.

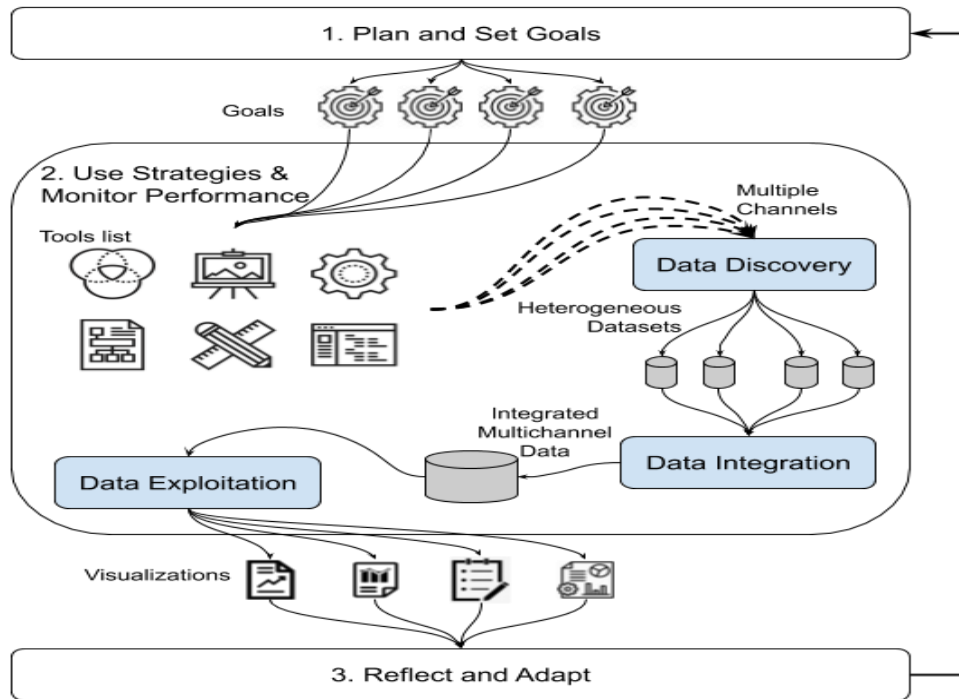


Figure 1: Self-Regulated Learning and use of a Data Value Chain to process multichannel data

3 ISSUES IN MULTICHANNEL DATA DISCOVERY AND INTEGRATION

There are challenges in each of the data processing activities (from data collection to decision making) under three groups (data discovery, integration and exploitation) of the DVC in order to process multichannel data in the context of SRL. In this positional paper, we have just highlighted (see Table 1) on the challenges associated with data discovery and integration.

Table 1: Challenges associated with data discovery, and data integration processing activities in the context of SRL

Group	Data Processing Activity	Challenges
Data Discovery	Collect and Label	-How multiple channels are streamlined and synchronized? -How heterogeneous datasets are labelled? -How to use a common time synchronization protocol across various tools and platforms?
	Prepare	-What is the ground truth in order to prepare datasets (dealing with noises, and redundant values)? -What is the contextual information of the SRL situation that should be accounted to guide the processing of multichannel data?
	Organize	-How to map the multichannel data to metrics of learning progress or learning curve?

		-How to select the transformation and aggregation decisions for processing multichannel data? Who is responsible to make such decisions especially when SRL practitioners have limited data literacy in most of the cases?
Data Integration	Integrate	-What are the correlations among heterogeneous datasets for coherently processing the multichannel data? -How the multichannel data will be integrated – low level (timestamp-based) or high-level (events-based)? -What are the rules and decisions required in order to integrate heterogeneous datasets into one dataset? Who is going to provide? What is the source to fetch such information?

4 DISCUSSION POINTS

We propose to use a DVC in the context of SRL to monitor the performance of learners. In this paper, we highlight some of the challenges associated with data discovery and integration activities. We would like to discuss these challenges and the different approaches from the participants in more detail during the workshop. Moreover, we would also like to discuss about the different data processing approaches for educational data than commercial data.

REFERENCES

Use the JLA_BODY_REFERENCE style.

APA referencing style <http://guides.library.uwa.edu.au/c.php?g=324904&p=2809540>

- Zimmerman, B. J., & Tsikalas, K. E. (2005). Can computer-based learning environments (CBLEs) be used as self-regulatory tools to enhance learning?. *Educational Psychologist*, 40(4), 267-271.
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational psychologist*, 25(1), 3-17.
- Dabbagh, N., & Kitsantas, A. (2012). Personal Learning Environments, social media, and self-regulated learning: A natural formula for connecting formal and informal learning. *The Internet and higher education*, 15(1), 3-8.
- Blikstein, P. (2013, April). Multimodal learning analytics. In *Proceedings of the third international conference on learning analytics and knowledge* (pp. 102-106). ACM.
- Worsley, M. (2012, October). Multimodal learning analytics: enabling the future of learning through multimodal data analysis and interfaces. In *Proceedings of the 14th ACM international conference on Multimodal interaction* (pp. 353-356). ACM.
- Shankar, S. K., Prieto, L. P., Rodríguez-Triana, M. J., & Ruiz-Calleja, A. (2018, July). A Review of Multimodal Learning Analytics Architectures. In *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)* (pp. 212-214). IEEE.
- Shankar, S. K., Ruiz-Calleja, A., Serrano-Iglesias, S., Ortega-Arranz, A., Topali, P., & Martinez-Monés, A. (2019). A Data Value Chain to Model the Processing of Multimodal Evidence in Authentic Learning Scenarios.

Collecting and Integrating Multimodal Data from a Programming Exercise Environment

Yuta Taniguchi

Kyushu University

taniguchi@ait.kyushu-u.ac.jp

Atsushi Shimada

Kyushu University

atsushi@ait.kyushu-u.ac.jp

ABSTRACT: Learning programming languages is one of the tough problems for students. This paper proposes four types of learning activities found in programming exercise contexts. We discuss their usefulness and how to collect such activities as data. Furthermore, we show two examples of analysis based on the data.

Keywords: Programming exercise, learning activity, e-textbooks

1 INTRODUCTION

Computer programming has been an essential skill for engineers and computer scientists for the past decades. Generally speaking, programming education for novice students are considered hard. For example, in C programming course, it is tough for beginners to write even a compilable source code from scratch, and they tend to write a program that will result in compile errors. However, since compile error messages are not always straightforward and beginner students do not have enough experiences to interpret error messages, fixing such errors takes a long time. Although such learning experiences could lower their motivation, not all students ask their teachers or friends when they have a trouble. Therefore, it is inevitable for teachers to actively intervene in programming exercise.

From the perspective of learning analytics, temporal records of students' learning process give plenty of insights to decide who and when we should support. For example, (Blikstein, 2011) collected logs of students' actions and analyzed coding strategies and temporal change of code, (Fu, Shimada, Ogata, Taniguchi & Suehiro, 2017) presented a dashboard system that visualize current situations of students' progress in programming exercise in real-time fashion. However, most of the existing studies only focused on coding activities.

Students' ability to find information necessary for making progress on their exercise has been rarely considered. For example, if we want to know whether a student will need a help or not, evaluating such abilities would be important as well as the ability to solve errors. Thus, we need novel approaches that take into account the information acquisition processes such as referring external learning materials, and/or asking classmates/teachers. Collecting such activity logs and integrating them with exercise activity logs would enable us to know students' situations in detail and support them in much more effective way.

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

2 THE LEARNING ACTIVITY TYPES IN EXERCISE

Learning activities previously studied are limited. Considering where students acquire the necessary information or knowledge during programming exercise in face-to-face classrooms, we propose four learning activity types students do during exercise. As shown in Figure 1, there are four targets around a student with which a student possibly interacts as a part of exercise activities.

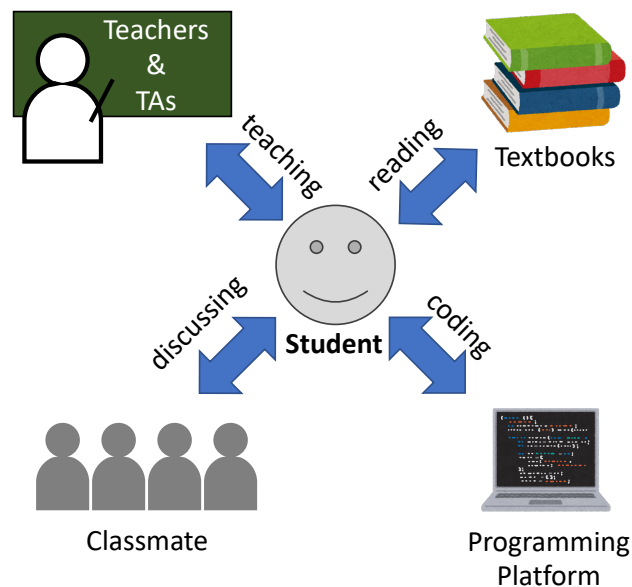


Figure 1. Four types of learning activities could be observed in programming exercise environments.

2.1 Interactions with Teachers and Teaching Assistants

Interacting with teachers and teaching assistants (TAs) are the representative activities found during exercise. Through lectures, teachers deliver important knowledge and information which would be necessary in exercise. Students usually follow a teacher and read the same page of a textbook as the teacher. In other situation, during exercise, for example, teachers and TAs would also directly interact with students to help them solve errors.

With e-textbook system, the former learning activity is relatively easy to record. For example, we can record which page a student is reading at a point of time. The latter type of interaction, however, needs additional mechanism since such interactions are usually done orally. The easiest way to collect such activities is asking teachers and TAs to record the names of students they interacted with. It might also be possible to use microphone for recording dialogue. Introducing a chat system for question answering would be another option that could be realized relatively easily.

2.2 Reading Textbooks

Students also read textbooks autonomously. For instance, during exercise, a student may consult a textbook for checking a grammar of a programming language to fix compilation errors. Even though

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

textbooks do not always provide information required to fix errors, they are definitely one of the authorities for students who do not have enough debug experiences. In addition, highlighting or note-taking are the other essential learning activities.

The usage of textbooks can be recorded with an e-book viewer system. Such a system is usually capable of storing users' activity logs into a database. Basically, page flipping actions and annotating actions can be collected. For more detailed reading behavior some researchers used eye trackers, but usually such devices cost so much that we cannot deploy at scale.

2.3 Coding Activities

As many literatures focused on, coding activity is a principal learning activity in programming exercise. Recording such activity tells us how a student worked on assignments. Monitoring students' activity in real-time give us chances for intervention to help students in trouble.

Usually a kind of educational programming platform is employed to record a student's actions, for example keyboard inputs, source code, and compilation results.

2.4 Discussion with Classmates

Classmates also play an important role in exercise. Sometimes, experienced teachers or TAs cannot understand why a novice programmer does not understand concepts, and the difference of knowledge could be a barrier in the communication between teachers and students. Usually students have a relatively similar level of understanding, and it could be considered that the communication between students is much easier. Actually, we can easily find students having a talk with each other during exercise instead of asking teachers or TAs.

Since such discussions may include sensitive information and happen everywhere in a class, to collect this kind of activity logs would be difficult ethically and technically. One feasible solution could be to introduce online forums where every post is public and optionally participants are anonymized. This idea could have potential to reduce the risk of ethical problem and make it easy to collect data. However, there is still a problem that we need to have student actively use forums.

3 POSSIBLE ANALYSIS

In this paper, we focus on the modalities of reading textbooks and coding activities. Figure 2 shows an example of visualized learning processes consists of them for each student (a horizontal line), which is composed of reading events (yellow dots), successful compiling events (green dots), and failure compilation events (red dots). We can see diversity of exercise processes in the picture. The visualization would be helpful for teachers to understand the current situation of a class, and temporal analysis on the processes has potential to predict the adequate timing of intervention.

Figure 3 shows an example of analysis on such learning processes from our previous study (Taniguchi, Okubo, Shimada & Konomi, 2018). The picture shows the aggregated contributions of each textbook page to resolve each error. Pages are considered positively contributed if a student solved the error message after reading the page and so forth. This kind of analysis enable us to find the problem of

textbooks or to identify students who cannot solve compile errors which are poorly described in textbooks.

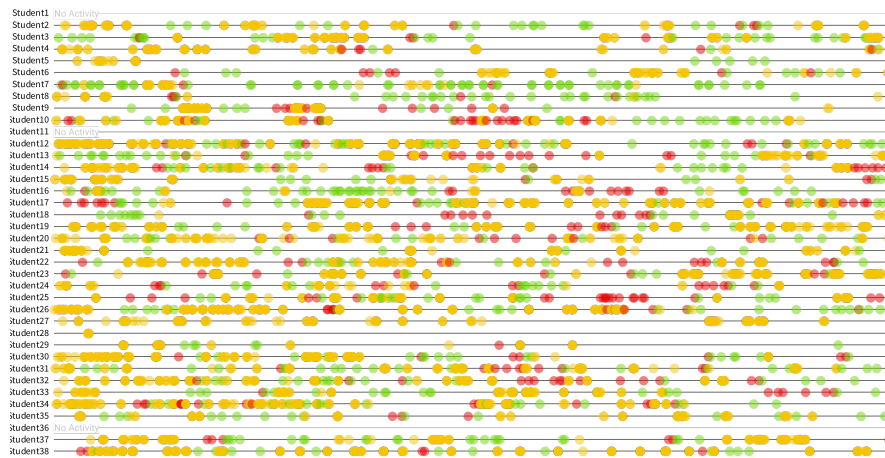


Figure 2. An example of learning processes of programming exercise consisting of e-book events and compilation events.

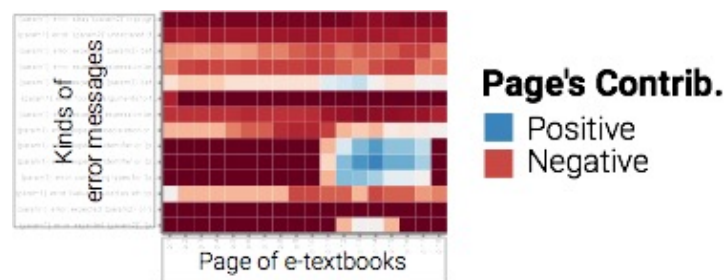


Figure 3. The result of page-error relation analysis. The positive or negative contribution of reading a specific page to solve a specific error is shown.

4 CONCLUSION

We introduced four types of students' learning activities and discussed possible usage and the way to collect. Our consideration mainly focused on how to collect learning activity data through Web-based system though. We are looking forward to discussing other types of activities and the way to collect learning activity logs from educational context not limited to programming environment.

REFERENCES

- Blikstein, P. (2011). *Using Learning Analytics to Assess Students' Behavior in Open-ended Programming Tasks*. Paper presented at the First International Conference on Learning Analytics and Knowledge.
- Fu, X., Shimada, A., Ogata, H., Taniguchi, Y., & Suehiro, D. (2017). *Real-time Learning Analytics for C Programming Language Courses*. Paper presented at the Seventh International Conference on Learning Analytics and Knowledge.
- Taniguchi, Y., Okubo, F., Shimada, A., & Konomi, S. (2018). *How are Students Struggling in Programming? Understanding Learning Processes from Multiple Learning Logs*. Paper presented at the 11th International Conference on Educational Data Mining.

EdRecSys Workshop@LAK2020

**Martin Hlosta¹, Christopher Krauss², Katrien Verbert³, Geoffray Bonnin⁴,
Martijn Millecamp³, Vaclav Bayer¹**

The Open University¹, Fraunhofer FOKUS², Berlin, Katholieke Universiteit Leuven³,
Université de Lorraine⁴

martin.hlosta@open.ac.uk, christopher.krauss@fokus.fraunhofer.de

ABSTRACT: The aim of the Workshop is to bring together researchers and practitioners that are working on topics related to the design, development and evaluation of recommender systems in educational settings as well as present the current status of research in this area and create cross-disciplinary liaisons between the RecSys and LAK communities. Overall, it aims to outline the rich potential of LAK as an application area for recommender systems, as well as expose participants to the challenges of developing such systems in a learning analytics context.

Keywords: Educational Recommender Systems, Learning Dashboards, Recommender Systems Evaluation, Adaptive curriculum, Learning Analytics

1 ORGANIZERS & PROGRAM COMMITTEE

1.1 Organizers

1. Christopher Krauss - Fraunhofer FOKUS, Berlin, Germany
2. Martin Hlosta - Knowledge Media Institute, The Open University, UK
3. Katrien Verbert - KU Leuven, Belgium
4. Martijn Millecamp - KU Leuven, Belgium
5. Geoffray Bonnin - Université de Lorraine, France
6. Vaclav Bayer - KMi, The Open University

1.2 Program Committee

1. Agathe Merceron - Beuth University of Applied Sciences
2. Zdenek Zdrahal - Czech Institute of Informatics, Robotics, and Cybernetics
3. Anne Boyer - anne.boyer@loria.fr
4. Armelle Brun - armelle.brun@loria.fr
5. Azim Roussanaly - azim.roussanaly@loria.fr
6. Hendrik Drachsler - drachsler@dipf.de
7. Rwitajit Majumdar - rwitajit@gmail.com
8. Brendan Flanagan - flanagan.brendanjohn.4n@kyoto-u.ac.jp
9. Dietmar Jannach - dietmar.jannach@aau.at
10. Christothea Herodotou, Institute of Education, The Open University, UK
11. Marek Hatala, Simon Fraser University, Canada
12. (+organisers)

2 BACKGROUND

Recommendation methods, techniques and systems open an interesting new approach to facilitate and support learning and teaching. There are plenty of resources available on the Web, both in terms of digital learning content and people resources (e.g. other learners, experts, tutors) that can be used to facilitate teaching and learning tasks. Recommendation methods have been used by the Educational Technology community to help identify suitable learning resources from a potentially overwhelming variety of choices. In addition, recommendation techniques are being used in learning analytics dashboards to provide actionable feedback to learners based on observed behavior and to recommend peer learners.

The aim of the Workshop is to bring together researchers and practitioners that are working on topics related to the design, development and evaluation of recommender systems in educational settings as well as present the current status of research in this area and create cross-disciplinary liaisons between the RecSys and LAK communities. Overall, it aims to outline the rich potential of LAK as an application area for recommender systems, as well as expose participants to the challenges of developing such systems in a learning analytics context.

Moreover, the Workshop aims to be an interactive, engaging experience that will motivate participants to get involved and start fruitful discussions on its topics. For that, it will combine several activities. On the one hand, a highly recognized keynote speaker will be invited to open the workshop. On the other hand, the Workshop would like to give to participants the opportunity to be engaged into creative and motivating discussions about the key issues related to LAK recommender systems. To this end, a panel of selected experts will be asked to pose a number of key questions that are related to enablers and challenges for recommender systems for learning, and then facilitate the discussion of these questions in a number of dedicated Working Groups. Topics of interest are, for instance, to join forces for the creation of a datasets challenge, share open educational datasets or best practices for the evaluation of recommender systems in a reliable way. Each Working Group will be expected to work around one particular question, and then report back to the plenary about the highlights of their discussion. In this way, all participants will be involved in an interactive discussion process and contribute to the joint outcome of the Workshop.

Moreover, papers submitted to the workshop and accepted by the Program Committee will be presented during the workshop. However, the presentations are not given by the authors themselves. Instead, accepted papers will be presented this time by other authors in 5-10 minutes each. Thus, each author has to deal with a different topic in advance and so the workshop becomes more interactive. Subsequently, the actual author has time to comment briefly and to supplement explanations. After each presentation there will be a short discussion with all workshop participants about the paper. The Workshop is expected to end with a small ceremony for giving the best paper awards.

The workshop builds upon the earlier EdRecSys workshop series that were organized at UMAP 2017 and WI'16; and the RecSysTEL workshop series that were organized at EC-TEL in 2012 and 2010.

2.1 TOPICS

1. Algorithms and design concepts for educational recommender systems
2. Recommender techniques for generating actionable feedback in learning dashboards
3. Study Material and (open) course recommendations
4. Recommendations to facilitate learning
5. Effects of recommendations and addressing different learning types
6. User centric perspective of the recommenders, perception of students/teachers
7. Adaptive curriculum and course sequencing
8. Job recommendations & adaptive educational biographies
9. Matchmaking and recommendation of learning peers
10. Recommendations for informal learning & recommendations for social/ peer learning
11. Recommendations for life-long learning
12. Situation and time sensitive recommendations
13. Utilizing the content/full-text information for the study material recommendation
14. Explanations and transparency in educational recommender systems
15. Tools, specifications and standards for enabling educational recommender systems
16. Evaluation of the educational recommenders

EVENT DETAILS

Type of event: Mini-tracks/Symposia

Proposed schedule and duration: Half-day (expected 3.5 hours based on LAK19 workshops duration)

- 9:00 Welcome - 10 minutes
- 9:10 Keynote - 30 minutes
- 9:40 Presentation of the papers (7 submissions + small break)
 - similarly to VISLA workshop at LAK19, we plan that the presentation will be done both by the authors as well as some other participant presenting the author's paper.
- 11:30 Group activity
 - Panel discussion, and possible discussions to join forces to create a dataset challenge (for next LAK) or to share open data sets
- 12:30 End of workshop

Type of participation: 'mixed participation'

The workshop/tutorial activities that participants should expect: discussion groups, presentations

Expected participant numbers: 15-20 participant

Planned dissemination activities to recruit attendants: distributing in the EDM/db-world/AI*A mailing lists, LAK google group, placing in the wikicfp, tweeting with the hashtags: #EdRecSys20, #EducationalRecommender, #RecommenderSystems, #LAK20, #LearningAnalytics, #EducationalDataMining

Additional required equipment for the workshop:

- pinboard(s) 1 piece
- flipchart(s) 2 pieces (10 sheets each)
- Post-its 5 piece(s) / packs
- different colors
- seating : u-shape
- apple adapter

3 OBJECTIVES AND INTENDED OUTCOMES

We aim to create links between the Educational Data Mining, Recommender Systems and LAK communities, not only by sharing research papers of these communities but also by initiating a panel discussion between all the participants, and possibly also joining forces in creating datasets for future EdRecSys editions. We expect 3-5 publications that would be published within “LAK Companion Proceedings”. Before the start and during the workshop, we will tweet with hashtags #EdRecSys20.

4 WEBSITE

Main menu with four navigation points:

- General Information (Motivation, Objectives, Workshop history)
- Submissions (Submission details, publication, important dates)
- Organisation (Workshop chairs, program committee)
- Accepted Papers and Schedule (schedule, accepted papers)

Important dates

- 29 October 2019: Workshop calls for participation announced
- 15 December 2019: Workshop papers submission deadline
- 9 January 2020: Notifications sent out (prior to early-bird registration deadline of 20 January 2019)
- 7 February 2020: Final version of paper due for LAK Companion Proceedings

Find the Optimizing Time intervals in Programming Learning for College Students

Fangjing Ning

Faculty of Education, Beijing Normal University
1173773581@qq.com

Baoping Li*

Faculty of Education, Beijing Normal University
libp@bnu.edu.cn

Penghe Chen

Beijing Advanced Innovation Center for Future Education, Beijing Normal University
chenpenghe@bnu.edu.cn

Qiuyu Chen

Faculty of Education, Beijing Normal University
15850586823@163.com

Xianru Zhang

Faculty of Education, Beijing Normal University
15668131800@163.com

ABSTRACT: Distributed learning is a learning strategy describing the allocation of learning time interval, which has an important impact on education. This study aims to find the optimizing learning time intervals in programming Learning. In a C programming course for college students, this study constructs the distributed learning Hidden Markov Model (HMM) based on the dataset collected from an Application named Dquiz. By learning with the application during one semester, the experimental results show that the AUC value is above 0.6, which proves that the HMM can be applied to the modeling of distributed learning. The study found that conceptual knowledge should be practiced in a medium interval of 4-7 days, procedural knowledge should be practiced in a short interval of 0-3 days, and comprehensive knowledge should be practiced in a long interval of more than 7 days.

Keywords: Hidden Markov Model, The Priming Effect of Programming Teaching, Contextual Variability, Learning time interval

1 INTRODUCTION

Distributed learning refers to students completing their learning tasks at several learning times (Griffin, 2016). The theories of priming effect and contextual variability have a deeper explanation for learning time interval. When the learning content is re-presented again in a short time, the learner can directly start the learning content by performing a small amount of processing because the learning content has been activated at the first presentation. However, this priming effect will gradually weaken with the increase of the interval between two presentations of the learning content. (Challis, 1993). In addition, the learning situation changes, which will increase characteristic memory coding related to the learning situation. Therefore, the learning effect increases with it, and then declines after reaching a certain peak, following a reverse U-shaped development curve

(Pashler, 2007). Learning time interval is an important aspect of learning process modeling. Nowadays, Hidden Markov Model (HMM) is the main state transition model and has been explored to identify the level of students (Leyzberg, Ramachandran, & Scassellati, 2018).

Students usually adopt the learning strategy of concentrated practice at the end of the term to deal with the examination, resulting in a shallow grasp of programming knowledge. Therefore, this study proposes a more efficient model based on hidden Markov model (HMM) to find the optimal learning time interval to improve learning efficiency and guide teachers how to arrange learning time according to knowledge in a course.

2 OPTIMIZING TIME INTERVAL STRATEGY WITH HIDDEN MARKOV MODEL

A distributed learning HMM is constructed from five aspects: the hidden state, observed state, the initial state distribution, the state transition probability distribution and the emission probability.

(1) The hidden state: mastered the knowledge or un-mastered knowledge. This study classifies the subject content into different knowledge types, such as conceptual knowledge, procedural knowledge, and comprehensive knowledge. (2) Observed state: the combination of learning time interval and answer result redefines students' observed state. The learning time interval can be further divided into a short interval (S, 0-3 days), a medium interval (M, 4-7 days) and a long interval (L, >8 days). Therefore, students' observation states include six states: a short interval-question right (S-R), short interval-question wrong (S-W), medium interval-question right (M-R), medium interval-question wrong (M-W), long interval-question right (L-R), long interval-question wrong (L-W). (3) The initial state distribution: before teaching, $P(L_0)$ denotes the initial probability of students' mastery of knowledge. (4) The state transition probability distribution: $P(L)$ is the probability of transfer from students' un-mastered knowledge to mastered knowledge between two exercises. (5) The emission probability: this study is based on the personalized setting of guessing probability and restarting failure probability of learning time interval. Guessing probability $P(G)$ is the probability that students can correctly answer questions because of the connection between previous learning knowledge and cognitive transfer when they have not mastered knowledge. Restart-failed probability $P(R)$ is the probability that students can't solve the problem correctly because of the failure of memory restart even if they master knowledge. For each time interval, there is a probability of guess $P(G|T)$ and a probability of restart-failed $P(R|T)$ where $T=\{s, m, l\}$.

3 EXPERIMENT AND EVALUATION

A C language mobile learning system have been developed, which allows students to log in at any time to do their homework. The system was applied in the C programming course of freshman in a university in Beijing in 2018. The course lasts 4 months and includes 255 exercises in total. Seventy-four students have completed all the questions, including 27 boys and 47 girls. The study selected students who tried to practice at different learning time intervals, as shown in Table 1. Eighty percent of the students in the data set were used to train the model, and 20% of the students were used to evaluate the model.

Table 1: Datasets and AUC.

	Number of Students	AUC
--	--------------------	-----

Conceptual knowledge	42	0.6209
Procedural knowledge	61	0.6016
Comprehensive knowledge	39	0.6912

Then this study re-coded the students' observation sequence as the correct or wrong answer (0 or 1), and evaluated the model with AUC value, as shown in Table 1. The model evaluation of three types knowledge's AUC is between (0.6, 0.7), so the model is acceptable.

Based on the well-trained decentralized learning HMM, the correct rate of students is calculated according to $P(G|T)$ and $P(R|T)$ at different learning time intervals. The optimal learning time interval has higher accuracy. It can be seen from the prediction of the correct rate of the answers in different intervals $P(\text{Right}|T)$ in Table 2 that the optimal learning time interval of the conceptual knowledge is a medium interval (i.e., 4-7 days). The optimum learning time interval of programming knowledge is a short interval (i.e., 0-3 days). The optimal learning time interval of comprehensive knowledge is more than 7 days.

Table 2: Accuracy at Different Time Intervals.

	$P(\text{Right} T=s)$	$P(\text{Right} T=m)$	$P(\text{Right} T=l)$
Conceptual knowledge	0.6251	0.9171	0.6666
Procedural knowledge	0.9666	0.1098	0.4914
Comprehensive knowledge	0.128	0.6622	0.9937

4 CONCLUSION

In conclusion, this study constructs a distributed learning model by incorporating the learning time interval into the observation sequence to explore the optimal interval. Data validation shows that the AUC indicators of three kinds of knowledge perform well in C programming course. The study found that conceptual knowledge is suitable for a medium interval, procedural knowledge for a short interval and Comprehensive knowledge for a long interval. This study solves the problem of the optimal learning time interval of different knowledge in program language learning and supports the implementation of mixed learning time strategy in the same course.

REFERENCES

- Challis, B. H. (1993). Spacing effects on cued-memory tests depend on level of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(2), 389-396.
- Leyzberg, D., Ramachandran, A., & Scassellati, B. (2018). The effect of personalization in longer-term robot tutoring. *ACM Transactions on Human-Robot Interaction (THRI)*, 7(3), 19.
- Pashler, R. H. (2007). Increasing retention without increasing study time. *Current Directions in Psychological Science*, 16(4), 183-186.
- Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257-286.

Mini Survival Kit: Prediction based recommender to help students escape their critical situation in online courses

Martin Hlosta, Vaclav Bayer, Zdenek Zdrahal

The Open University, United Kingdom

martin.hlosta@open.ac.uk

ABSTRACT: The poster focuses on a recommender method that is tightly related to predictive learning analytics in distance higher education focused on the identification of students at risk of not submitting their assignments and subsequently failing their courses. Given a lack of student time to the assignment deadline, the method aims to provide a minimalistic recommendation for students to increase their chances of submitting the assignment so that they survive a possible difficulty they encounter. We formally define the task as an optimisation problem and propose a simple algorithm that will serve as a baseline for further improvement. On an offline evaluation on one STEM course, taking only students predicted as at-risk, those that followed the recommendations were associated with higher submission rates than if they only accessed any online resource.

Keywords: Study Recommender, Recommender Systems, Predictive Modelling, Sensitivity Analysis, At-risk students.

1 INTRODUCTION AND RELATED WORK

The educational recommender systems (ERS) in a closed-course setting aim to guide learners to pass the course with the highest possible measured results by providing recommendations from a predefined number of materials or activities in the course. Compared to traditional recommender systems, the focus should be to make the learning more efficient, influence future learning activities, and assign high importance for the time when the activity is recommended (Krauss, 2018).

In distance higher education, student drop-out is still one of the main issues and identifying at-risk students followed by interventions is one of the ways of tackling the problem. A recommender utilising predictive modelling can help with subsequent recommendation of the remedial activities. However, this connection is usually omitted and the only research connecting these two we found is (Thai-Nghe et al., 2012). But rather than using the predictive analytics to guide the recommender, the authors used factorisation techniques used in recommender systems to predict student performance on unseen tasks in the tutoring systems.

We build on the results of a recommender strategy for closed-courses in (Huptych, Bohuslavec, Hlosta, & Zdrahal). The recommender focused on providing materials, where the student has fewer clicks compared to the successful students measured at the same time as the course in the previous run. The recommender lacks, however, how many materials should be recommended and doesn't take into account student's critical moment, i.e. when being at-risk of failing/not submitting the assignment. At this moment, the goal of the recommendation is to retain students and avoid drop-out.

3 METHODOLOGY

We aim to fill this gap by developing a recommender tied to the predictive model, which identifies students at risk of not submitting their next assignment in distance higher education. This has shown to be an early proxy for the later failure in the whole course (Hlosta, Zdrahal, Bayer, & Herodotou, 2020). We assume that students' measured activity (clicks in VLE) can have a positive impact on their results. The main idea of the recommender could be summarised as: "Given student's current effort, which activities increase their chances of succeeding in the following assignment?"

Available Features - In each week w , a set of static and dynamic features exists for each student. The static features FS represent the information known at the start of the course, e.g. demographic data. The set of dynamic features $FD_ALL_w = \{FD_1, \dots, FD_w\}$ includes number of clicks in VLE from week 1 to week w . Each set FD_i in FD_ALL_w include the number of clicks in M activities, which are available to students. The final set of features including the static features is $F_w = \{FS, FD_ALL_w\}$.

Given the student predictions generated in week n , we want to recommend personalised activities in week $n + 1$. Because the machine learning models are generated from the legacy data, in week n models both for week n and week $n + 1$ exist. We expect that the model can provide a confidence of student s in week w of not submitting (NS) the next assignment, defined as: $PR_w(NS, s_w)$. The confidence is value in $[0; 1]$, the higher the value, the higher the chance of student being at risk. The activities to be recommended in week n for the next week $n + 1$ are FD_{n+1} . The activities FD_{n+1} are not known in week n but thanks to the predictive model for both weeks n and $n + 1$, we can use a **sensitivity analysis** of the model to simulate the predictions in week $n + 1$. We assign students possible values of the dynamic features in the next week, i.e. their clicks. Together with the information known in the current week n , we let the model predict their confidence $PR_{n+1}(NS, s_{n+1})$. Simulating different combinations of values in the future allows us to estimate, which actions in the future lead to higher chances of submitting the assignment. In fact, for at-risk students in week n , having $PR_n(NS, s_n) > 0.5$, we aim to increase their submission prediction in week $n + 1$ so that $PR_{n+1}(NS, s_{n+1}) < t$, typically $t = 0.5$. This is depicted in Figure 1.

For a student s in week n we define the task as a constrained optimisation problem: $\min(\text{sumClicks}(s_{n+1}, FD_{n+1}))$ subject to $PR_{n+1}(NS, s_{n+1}) < t$. We developed a simple algorithm which, for a given student, iteratively searches for an activity where a click will maximally decrease the prediction of being NS in the next week. The algorithm continues until the constraint is met.

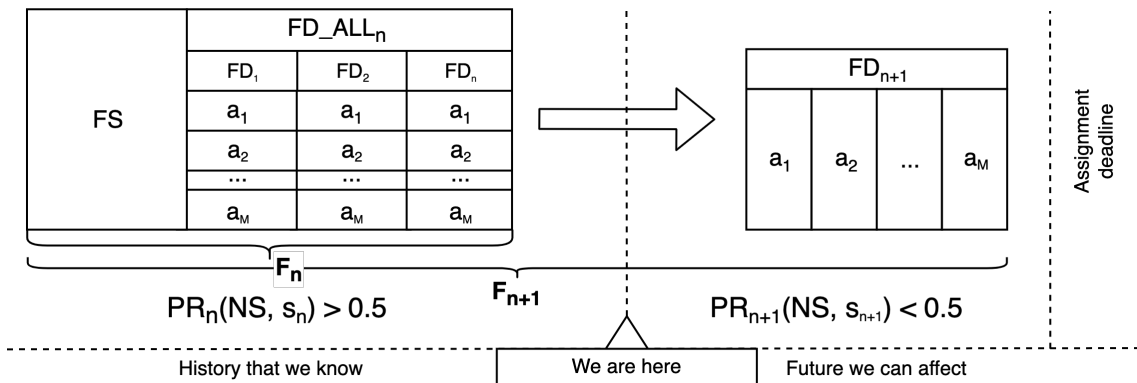


Figure 1 Schema of the recommender with one student predicted in week n as Not Submit.

4 EVALUATION AND RESULTS

We used data of 593 students in one STEM course focused on predicting submission of the first assignment (A1) with the deadline in week 9 with submission ratio 89%, i.e. predictions in weeks 1-8. For A1, 38 materials are available for students. The predictions with AUC of 0.81-0.96 were generated by the GBM model. We focused on week 7, i.e. two weeks before the deadline, where the correlation with the next week prediction was lowest, suggesting a change in students' behaviour which can be beneficial for the recommender.

To investigate the validity of the approach, we evaluated if students that followed the recommendation had a higher submission ratio than those who did not. We focused only on students predicted as NS, i.e. $PR_7(NS, s_7) > 0.5$ ($n=78$) and excluded students that had no click in VLE ($n=50$). Table 4 shows that 82% of predicted students did not submit (50% after filtering the non-active students). From the active students, if a student accessed all the recommended materials ($n=11$), the NS rate decreased to 36% compared to 58% if they did not follow the recommender (difference 22%).

Table 4 Comparison of the number of students following and not following the recommendations

		Followed the RS	All	Not Submit	Submit	rel(NS)
Total			78	64	14	0.82
Is active			28	14	14	0.50
Recommender all	YES		11	4	7	0.36
activities	NO		17	10	7	0.58

5 CONCLUSIONS AND FUTURE WORK

The comparison shows that following all the recommended activities is associated with higher chances of submitting for students that were predicted as not submit and active in VLE. Navigating students towards the submission, we are trying to disprove the predictions and lower the prediction performance, increasing False Positives. Students with increased last-minute activity showed as typical cases of False Positives (Hlostá et al., 2020). The results are promising, yet our next step will be investigating a higher number of students before deploying the recommender for online evaluation to teachers and students.

REFERENCES

- Hlostá, M., Zdrahal, Z., Bayer, V., & Herodotou, C. (2020). *Why Predictions of At-Risk Students Are Not 100% Accurate? Showing Patterns in False Positive and False Negative Predictions*. Paper presented at LAK'20, Frankfurt, Germany.
- Huptych, M., Bohuslavsek, M., Hlostá, M., & Zdrahal, Z. (2017). Measures for recommendations based on past students' activity.
- Krauss, C. (2018). Time-dependent recommender systems for the prediction of appropriate learning objects.
- Thai-Nghe, N., Drumond, L., Horváth, T., Krohn-Grimberghe, A., Nanopoulos, A., & Schmidt-Thieme, L. (2012). Factorization techniques for predicting student performance *Educational recommender systems and technologies: Practices and challenges* (pp. 129-153): IGI Global.

Culturally Inclusive Learning Analytics (#CILA)

Ioana Jivet^{1,2}, Tom Broos³, Maren Scheffel¹, Hendrik Drachsler^{4,5,1}

¹Open University of the Netherlands, ²TU Delft, ³KU Leuven,

⁴DIPF, ⁵Goethe University Frankfurt

ioana.jivet@ou.nl, maren.scheffel@ou.nl, tom.broos@kuleuven.be, drachsler@dipf.de

ABSTRACT: The Learning Analytics community will celebrate its 10-year anniversary at LAK20 in Frankfurt, Germany. In the past ten years, we have seen many examples of learning analytics being implemented in various countries. But we also increasingly realize that learning analytics are applied very differently in those countries. This makes the transfer of learning analytics solutions from one country to another rather difficult. Today, the implementation of learning analytics is mainly a social challenge rather than a technical one. Within the CILA workshop we want to work out cultural differences for the adoption of learning analytics within the international LA community. During the workshop, we will use Hofstede's framework as a starting point for analysis of several existing LA tools and build upon this activity toward a facilitated discussion on the cultural differences for the adoption and use of learning analytics. As input for the discussion, we will also present the outcomes of a cross-cultural survey that collects input on the acceptance and use of LA tools around the world, as well as the workshop participants' own answers to the survey.

Keywords: learning analytics, user-centered design, culture.

1 BACKGROUND

Within the Learning Analytics (LA) community, the idea that a “one size fits all” paradigm does not lead to effective LA tool designs has become widely accepted, but there is still a big question mark over what factors define the “right size” for every learner. During this workshop, we wish to explore whether the learners' cultural background and the institutional culture surrounding the learning context are some of these factors. In an increasingly international educational landscape, how far should the LA community go in taking such factors into account? What opportunities are offered by LA technologies to deal with cultural barriers and how can we design inclusive LA tools that minimise these barriers?

The field of LA has gained a lot of attention in the last years as more and more data about learners and their contexts became available. The number of LA tools implemented in online learning environments that bring together learners from all over the world has been steadily on the rise. At the same time, LA implementations are being transferred across institutions and even across countries and continents. For example, the LALA project (<https://www.lalaproject.org/>) is a European capacity building project that aims to improve the quality of Higher Education in Latin America, by enabling local universities to implement LA. Similarly, the Competen-SEA project (<http://competen-sea.eu/>) aims to enable universities in South-East Asia to develop a new kind of accessible, affordable, high quality and effective educational services in order to reach groups of the

population now excluded from traditional education by leveraging European experiences. The success of both projects relied on adapting the technology to the local context.

Student learning patterns and learning strategy use in higher education have been shown to differ across different cultures (Marambe, Vermunt & Boshuizen, 2012). Cultural differences play a significant role also in online learning influencing students' collaborative learning (Vatrapu & Suthers, 2007), as well as educational technology acceptance and use (Nistor, Göğüş & Lerche, 2013). Instructional designers and teachers need to build culturally inclusive learning course designs in order to encourage full participation by international students (Liu et al., 2010; Gómez-Rey et al., 2016). On an institutional level, managers of transnational higher education partnerships believe culture affects both the academic and operational management of the education programs (Eldridge & Cranston, 2009). While there were some initial attempts to focus learning analytics on cultural differences (Vatrapu, 2011), the topic is widely underrepresented in current LA work. Still, the few studies that included the cultural dimension in their research show that cultural differences influenced the effectiveness of LA interventions (Mittelmeier et al., 2016; Davis et al., 2017; Kizilcec & Cohen, 2017; Kizilcec et al., 2014), and that cultural differences can be detected based on analytics methods in learning technologies (Rüdian et al., 2019).

The LAK community would benefit from starting a discussion and drafting a set of suggestions on how to create more culturally inclusive tools that put users and their needs at the centre of the design process. Following such principles could lead to more meaningful tools that do not put certain stakeholder groups at an advantage over others. During the workshop, we will use Hofstede's framework as a starting point for analysis of several existing LA tools and build upon this activity toward a facilitated discussion. Although the framework is not free of criticism (Baskerville, 2003), it remains an accessible and highly popular model. There is extensive research grounded in this framework as to how culture influences different aspects of learning, including beliefs about learning and the educational practice (Hofstede, 1986). As input for the discussion, we will also present the outcomes of a cross-cultural survey that collects input on the acceptance and use of LA tools around the world.

2 ORGANISATIONAL DETAILS

2.1 Objectives

We would like to address the following objectives during the workshop:

1. Raise awareness on the effects of culture on learning and beliefs about learning and implications on: (i) the acceptance and use of LA tools, (ii) the design requirements of LA tools, (iii) the potential for reusability and the need for adaptation of LA tools across different cultural contexts, (iv) the need for adaptation of LA tools within cultural heterogeneous environments.
2. Introduce the cultural background of LA stakeholders on the list of design considerations in the LA domain both in practice and in research.

3. Draw an initial list of recommendations based on workshop participants' experience with respect to the integration (or lack thereof) of cultural aspects into the design, implementation, evaluation and use of LA tools.

2.2 Organisers

The workshop organisers have combined more than 20 years of experience in TEL, teaching, applying and LA in different cultural contexts. For the purpose of this workshop, we will be working with representatives of different cultures to gather their input.

2.3 Schedule and participants

We expect between 25-40 participants. Because a culturally heterogeneous audience would contribute to the discussions in the workshop, we aim at the representation of at least 7 different countries (at least 2 non-European). For this we will send a number of targeted invitations, complementing the open channels for workshop invitations. The CILA workshop will have a mixed participation design, in general, it is an open workshop but participants need to fill in survey documents in advance to collect some data about cultural differences of LA.

The workshop is planned for a full day. A rough schedule will be:

Welcome and initial remarks

Part I: Introduction to cultural background frameworks and research

- Invited talk on cultural differences by an external expert

Coffee Break

Part II: Groupwork

- Analysis of the design of LA tools from the participants' cultural perspective
- Roundtable groupwork outcomes discussion

Lunch

Part III: Survey outcomes - input from the survey on LA acceptance, adoption and use practices around the world collected for this workshop

Coffee Break

Part IV: Workshop outcomes

- Facilitated discussion based on several statements in a debate format
- Open discussion and drawing further action points

REFERENCES

Baskerville, R. F. (2003). Hofstede never studied culture. *Accounting, organizations and society*, 28(1), 1-14.

- Davis, D., Jivet, I., Kizilcec, R. F., Chen, G., Hauff, C., & Houben, G. J. (2017, March). Follow the successful crowd: raising MOOC completion rates through social comparison at scale. LAK'17 (pp. 454-463). ACM.
- Eldridge, K., & Cranston, N. (2009). Managing transnational education: does national culture really matter?. *Journal of Higher Education Policy and Management*, 31(1), 67-79.
- Gómez-Rey, P., Barbera, E., & Fernández-Navarro, F. (2016). The impact of cultural dimensions on online learning. *Journal of Educational Technology & Society*, 19(4), 225-238.
- Hofstede, G. (1986). Cultural differences in teaching and learning. *International Journal of intercultural relations*, 10(3), 301-320.
- Keller, J. M., Ucar, H., & Kumtepe, A. T. (2018). Culture and motivation in globalized open and distance learning spaces. In *Supporting Multiculturalism in Open and Distance Learning Spaces* (pp. 146-165). IGI Global.
- Kizilcec, R. F., Schneider, E., Cohen, G. L., & McFarland, D. A. (2014). Encouraging forum participation in online courses with collectivist, individualist and neutral motivational framings. EMOOCS 2014, Proceedings of the European MOOC stakeholder summit, 80-87.
- Kizilcec, R. F., & Cohen, G. L. (2017). Eight-minute self-regulation intervention raises educational attainment at scale in individualist but not collectivist cultures. *Proceedings of the National Academy of Sciences*, 114(17), 4348-4353.
- Liu, X., Liu, S., Lee, S., & Magjuka, R. J. (2010). Cultural differences in online learning: International student perceptions. *Educational Technology & Society*, 13(3), 177-188.
- Marambe, K. N., Vermunt, J. D., & Boshuizen, H. P. (2012). A cross-cultural comparison of student learning patterns in higher education. *Higher Education*, 64(3), 299-316.
- Mittelmeier, J., Tempelaar, D., Rienties, B., & Nguyen, Q. (2016). Learning Analytics to Understand Cultural Impacts on Technology Enhanced Learning. *International Association for Development of the Information Society*.
- Nistor, N., Göğüş, A., & Lerche, T. (2013). Educational technology acceptance across national and professional cultures: a European study. *Educational Technology Research and Development*, 61(4), 733-749.
- Rüdian, S., Vladova, G., Gundlach, J., Kazimzade, G., Pinkwart, N.: Predicting Culture and Personality in Online Courses. Proceedings of the first International Workshop on Supporting Lifelong Learning co-located with AIED 2019, Chicago, USA, June 29, 2019. CEUR Workshop Proceedings.
- Vatrapu, R. K. (2011, February). Cultural considerations in learning analytics. In LAK (pp. 127-133).
- Vatrapu, R., & Suthers, D. (2007, January). Culture and computers: A review of the concept of culture and implications for intercultural collaborative online learning. In *International Workshop on Intercultural Collaboration* (pp. 260-275). Springer, Berlin, Heidelberg.
- Yousef, A. M. F., Chatti, M. A., Schroeder, U., & Wosnitza, M. (2014, July). What drives a successful MOOC? An empirical examination of criteria to assure design quality of MOOCs. In *2014 IEEE 14th International Conference on Advanced Learning Technologies* (pp. 44-48). IEEE.

XLA: Explainable Learning Analytics

Katrien Verbert, Tinne De Laet, Martijn Millecamp, Tom Broos

KU Leuven, Belgium

{firstname.lastname}@kuleuven.be

Mohamed Amine Chatti, Arham Muslim

University of Duisburg-Essen, Germany

{mohamed.chatti, arham.muslim}@uni-due.de

ABSTRACT: The aim of learning analytics is to turn educational data into insights, decisions, and actions to improve learning and teaching. Many of the learning analytics applications behave often as a black-box. The reasoning of the provided insights, decisions, and actions are often not explained to the end-user, and this can lead to trust issues when interventions, feedback, and recommendations fail. This proposal describes the goal and activities of the **LAK 2020 full-day workshop** on the design and development of explainable learning analytics applications that **provide transparent insights** and **explain decisions and actions** to the intended end-users **with visualisations**. The main theme of the workshop is to explore how user interfaces that explain how decisions and actions were taken can strengthen the adoption of learning analytics applications. The key findings, recommendations, and conclusions of the workshop will be presented in a summarising report, highlighting best practices as well as future work opportunities.

Keywords: learning analytics dashboards, explanations, actionability

1 THEME AND WORKSHOP BACKGROUND

Currently, the Learning Analytics (LA) domain is maturing and often provides **insights** in and **decisions/actions** (e.g. interventions, feedback, recommendations) regarding learning and teaching behaviors. Before the insights, decisions, and actions can actually impact learning, the process and outcomes of the data analysis have to be **transparent** to the end-users. While advanced machine learning might create accurate insights, decisions, and actions, they will not be per se trusted by the user. Opening the black-box of LA to the user, in a user-tailored fashion is the first step towards providing transparent insights, decisions, and actions. Approaches for obtaining transparency are key to improve the trustworthiness, impact and adoption of LA systems at scale. Learning analytics dashboards are at the core of the LAK vision to support better decision-making [Verbert et al. 2020]. As coined by Erik Duval, the use of visualisation techniques in such dashboards is also key to support transparency [Duval 2011].

The evolution towards more transparent LA is urgent, as recent data protection and privacy regulations (EU GDPR and CCPA (California Consumer Privacy Act)) stipulate that transparency is a fundamental right. Even more, they state that each user has the right to withdraw him/herself from automatic decision making and profiling. So, it is now time to act and to jointly work towards transparent insights, decisions, and actions in the domain of LA.

The motivation of the workshop is to advance the research and practices around transparent LA. To this end, the workshop will bring together researchers, practitioners, educational developers and policymakers in an interactive workshop format. We welcome contributions of both long and short papers, and including both research and practitioner papers, around the general theme of explainable insights, decisions, and actions in LA including:

- Theories and methods for transparent insights, interventions, feedback, and recommendations in LA
- Explanations for transparency in LA
- Visualisations for transparency in LA
- User involvement/control for transparency in LA
- Explainable user/learner modeling
- Methods to assess transparency and explainability in LA
- Impact of explainable LA on the stakeholders (e.g. motivation, engagement and adoption)
- Case studies demonstrating the need for explainable LA

We organised a related edition of our workshop at EC-TEL 2019, and the workshop builds upon the earlier VISLA workshop series on visual approaches to learning analytics, organised at LAK 19 and LAK 15.

2 ORGANISATIONAL DETAILS

After the program committee makes a selection of accepted submissions, we will make them openly available before the start of the workshop.

During our 1-day workshop, we aim to facilitate a very interactive and engaging event where we want to avoid death by powerpoint at all causes and promote discussion activities over presentational ones. In the first half of the workshop, we will therefore ask participants to shortly present the work of *another* submission and to relate it back to their *own* work. The facilitators will allocate challengers per presentation to move the discussion around common themes and differences in approaches.

3 ABOUT THE ORGANISERS

Katrien Verbert is Associate Professor of the HCI research group of the Department of Computer Science at KU Leuven, Belgium. Her research interests include recommender systems, visualization techniques, visual analytics, and applications in healthcare, learning analytics, precision agriculture and digital humanities.

Tinne De Laet is Associate professor at the Faculty of Engineering Science, KU Leuven. She is the Head of the Tutorial Services of Engineering Science. Her research focuses on using learning analytics, conceptual learning in mechanics, multiple-choice tests, and study success.

Tom Broos is a PhD student at the HCI research group of KU Leuven, Belgium. He researches scalable learning analytics interventions to support first-year students in their transition to higher education. He emphasizes the active receiver, analytical transparency and privacy.

Martijn Millecamp is a PhD student at the HCI research group of KU Leuven, Belgium. His research interests include user interfaces for music recommender systems and dashboards for learning analytics.

Mohamed Amine Chatti is a Professor of Computer Science and head of the Social Computing Group in the Department of Computer Science and Applied Cognitive Science at the University of Duisburg-Essen, Germany. His research interests include data science, learning technologies, and social computing.

Arham Muslim is a senior researcher in the Social Computing Group, Department of Computer Science and Applied Cognitive Science at the University of Duisburg-Essen, Germany. His research focuses on open learning analytics, information visualization, and visual analytics.

ACKNOWLEDGEMENTS

The authors also gratefully acknowledge the support of the European Commission through the LALA (Capacity to Use Learning Analytics to Improve Higher Education in Latin America) project Erasmus+ capacity building project (grant no. 586120-EPP-1-2017-1-ES-EPPKA2-CBHE-JP). The European Commission support for the production of this publication does not constitute an endorsement of the contents which reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

REFERENCES

- Duval, E. (2011). Attention please!: learning analytics for visualization and recommendation. *LAK*, 11, 9-17.
- Verbert, K., Ochoa, X., De Croon, R., Dourado, R. A., De Laet, T. (2020). Learning analytics dashboards: the past, the present and the future. In *LAK '20: International Learning Analytics and Knowledge (LAK) Conference*, March 23–27, 2020, Frankfurt, Germany. ACM, New York, NY, USA, 6 pages.

A Review of Explanatory Visualizations in Recommender Systems

Mouadh Guesmi, Mohamed Amine Chatti, Arham Muslim

Social Computing, Department of Computer Science and Applied Cognitive Science,
University of Duisburg-Essen, Germany

mouadh.guesmi@stud.uni-due.de, mohamed.chatti@uni-due.de, arham.muslim@uni-due.de

ABSTRACT: A key requirement for the acceptance and adoption of learning analytics is trust. Explanation facilities can provide means to achieve transparent and trustworthy learning analytics. A possible way to implement explainable learning analytics (XLA) is to provide explanations in educational recommender systems. Recently, explainable recommendation has attracted much attention in the recommendation research community, but remained under-explored in the domain of learning analytics. In this paper, we systematically review the work on explanations in the recommender system literature. Recognizing the importance of visualization in learning analytics research and practice, we approach the recommender system literature from the angle of explanatory visualizations, that is using visualizations as a display style of explanation. The aim of this review is to provide a systematic overview on the current design of explanatory visualizations in the explainable recommendation literature in order to inspire learning analytics researchers and practitioners to investigate explanatory visualizations as a means of establishing transparency and trust in educational recommender systems.

Keywords: Explainable Learning Analytics, Explainable Recommendation, Visualization, Explainability, Transparency, Trust.

1 INTRODUCTION

Trust has recently gained much attention in the learning analytics community (Drachsler & Greller, 2016; Pardo & Siemens, 2014). Providing explanations for system-generated decisions and actions (e.g. recommendation, intervention, feedback) has the potential to make learning analytics transparent and trustworthy. Explainable learning analytics (XLA) refers to opening the learning analytics black-box to the user in order to provide transparent insights, decisions, and actions. This is key to improve the trustworthiness, acceptance, and adoption of learning analytics systems at scale. Explainable recommendation represents an important branch of XLA. While there has been a considerable amount of research on recommender systems in the domain of learning analytics (Verbert et al., 2012), providing explanations of the outputs of these systems is under-investigated. Visualization techniques that are widely used in learning analytics dashboards can provide a visual entry to explain the output of recommender systems. Thus, it becomes important to develop a better understanding on the potential of visually explainable recommendation to increase the transparency and trustworthiness of learning analytics. The primary aim of this work is to provide learning analytics researchers and practitioners with an overview on the visual explanation design approaches used in the recommendation research community in order to inspire them to adopt and adapt this work in the learning analytics domain and leverage explanatory visualizations to establish transparency and trust in educational recommender systems.

In this paper, we systematically review the work on explainable recommendation that has been published in the last decade. We conduct a comprehensive survey of explainable recommendation with a focus on explanatory visualizations, that is using visualizations as a display style of explanation. Various surveys of explainable recommendation research with many classification taxonomies were proposed in the literature (Friedrich & Zanker, 2011; Gedikli, Jannach, & Ge, 2014; Nunes & Jannach, 2017; Papadimitriou, Symeonidis, & Manolopoulos, 2012; Zhang & Chen, 2018). However, a classification of the literature based on visualization as a display style of explanation is lacking. With this paper, we aim to fill that gap.

The rest of this paper is organized as follows. Section 2 gives an overview on explainable recommendation concepts with a focus on the explanation display. Munzner’s what-why-how visualization framework is briefly explained in Section 3. This framework provides the theoretical background to discuss how explanatory visualizations are provided in the reviewed literature. Section 4 presents the results of this survey discussed based on Munzner’s framework. Section 5 provides a discussion of the obtained results. Section 6 finally summarizes the main findings of this work.

2 EXPLAINABLE RECOMMENDATION

While for a long time research in recommender systems focused primarily on algorithmic accuracy, there has been an increased interest in more user-centered evaluation metrics such as transparency and trustworthiness of a recommender system (Knijnenburg, Willemsen, Gantner, Soncu, & Newell, 2012; Konstan & Riedl, 2012; Pu, Chen, & Hu, 2011, 2012). A major approach to enhance transparency and trust in recommender systems is to provide the rationale behind a recommendation in the form of explanations (Tintarev & Masthoff, 2015). Generally, explanations seek to show how a recommended item relates to a user’s preferences (Vig, Sen, & Riedl, 2009). The explainability of recommendations has attracted considerable attention in recent years. Explainable recommendation refers to personalized recommendation algorithms that not only provide the user with the recommendations, but also provide explanations to make the user aware of why such items are recommended. Explainable recommendation research covers a wide range of techniques and algorithms, and can be realized in many different ways (Tintarev & Masthoff, 2015). Recommendation explanations can also be presented in very different display styles, which could be a relevant user or item, an image, a sentence, a chart, etc. (Zhang & Chen, 2018). The display styles of the recommendation explanations can be generally classified into textual explanations and visual explanations. Textual explanations generate a piece of text information as recommendation explanation. To take advantage of the intuition of visual images, visual explanations provide the user with a visualization as explanation. The visualization can be a chart, an image (whole image or particular visual highlights in the image), or a graph, especially in social network related application scenarios.

In our survey, we focus on how visualizations can be leveraged to provide explanations in recommender systems. We use Munzner’s what-why-how visualization framework (Munzner, 2014) as a theoretical background to systematically discuss the visual explanation display approaches used in the reviewed literature. This framework is briefly explained in the next section.

3 THE WHAT-WHY-HOW VISUALIZATION FRAMEWORK

Munzner (2014) proposed an analysis framework of breaking down visualization design according to *what-why-how* questions that have *data-task-idiom* answers. The what-why-how visualization framework analyzes visualization use according to three questions: *what* data the user sees, *why* the user intends to use a visualization tool, and *how* the visual encoding and interaction idioms are constructed in terms of design choices. Each three-fold *what-why-how* question has a corresponding *data-task-idiom* answer.

3.1 What: Data Abstraction

The what dimension refers to the abstract types of *what* data can be visualized. The four basic dataset types are *tables*, *networks*, *fields*, and *geometry*; other possible collections of items include *clusters*, *sets*, and *lists*. These datasets are made up of different combinations of the five data types: *items*, *attributes*, *links*, *positions*, and *grids*. The type of an attribute can be categorical or ordered, with a further split into ordinal and quantitative (Munzner, 2014).

3.2 Why: Task Abstraction

The why dimension refers to tasks expressing the reason *why* a visualization tool is being used. A task is defined as an *{action, target}* pair where action is a verb defining use goals and target is a noun referring to some aspect of the data that is of interest to the user. In general, visualization query actions can have three scopes: *identify* one target, *compare* some targets, or *summarize* all targets. Targets for all kinds of data are *trends* (e.g. increases, decreases, peaks), *outliers*, and *features*, i.e. any particular task-dependent structures of interest (e.g. popularity, clusters, relationships). For one attribute, the target can be the *extremes*, i.e. minimum and maximum values, or the *distribution* of all values for an attribute. For multiple attributes, the target can be *dependency*, *correlation*, or *similarity* between them. The target with network data can be *topology* (i.e. structure of the network) in general or *paths* in particular, and with spatial data, the target can be *shape* (Munzner, 2014). Examples of task abstraction defined as *{action, target}* pairs include *{identify, trends}* or *{identify, outliers}* in all data; *{identify, extremes}* of one attribute; *{compare, correlation}* of multiple attributes; *{summarize, distribution}* of all attributes; and *{identify, paths}* in a network.

3.3 How: Idioms

The how dimension refers to *how* a visualization idiom can be constructed out of a set of design choices. These choices can be broken down into two major classes: (a) how to *encode* a visualization which includes how to arrange data spatially and how to map data with all of the nonspatial visual channels such as color, size, angle, and shape and (b) how to *interact* with a visualization. Interactions include how to *manipulate* a view e.g. change any aspect of the view, select elements from within the view, or navigate to change the viewpoint within the view; how to *facet* data between views e.g. by juxtaposing and coordinating multiple views, partitioning data between views, or superimposing layers on top of each other; and how to *reduce* the data by filtering data away, aggregating many data elements together, or embedding focus and context information together within a single view (Munzner, 2014).

4 SURVEY

In this survey, we classify existing explainable recommendation research in a systematic manner, based on the what-why-how visualization framework. Recognizing the importance of visualization in learning analytics, we focus on explanatory visualizations. Further, as learning analytics is a data-driven approach, we mainly focus on works that leverage data visualization for explainable recommendation. We surveyed 21 explainable recommendation tools which use visualizations as a display style of explanation. Some of these tools had more than one explanatory visualization. The set of selected tools was compiled based on recent reviews and works on explainable recommendation (He, Parra, & Verbert, 2016; Kouki, Schaffer, Pujara, O'Donovan, & Getoor, 2019; Millecamp, Htun, Conati, & Verbert, 2019; Tintarev & Masthoff, 2015; Zhang & Chen, 2018). The survey results are summarized in Table 1 and discussed in detail in the following sections.

Table 1: Summary of the survey results.

Tools	What (dataset)	Why	How		
			Encode	Interact	
		(identify, distribution)			No interaction
		(summarize, features)			Manipulate: Navigate
		(identify, features)			Manipulate: Change
		(summarize, similarity)			Facet: Juxtapose
		(compare, similarity)	+	+	Reduce: Filter
		(identify, topology)			Manipulate: Select
		(identify, paths)	+	+	
Relevance Tuner+ (Tsai & Brusilovsky, 2019)	Geometry	+			
Hyper (Kouki et al., 2019)	Sets	+	+		
Tagsplanation (Vig et al., 2009)	networks	+			
PeerChooser (O’ Donovan et al., 2008)	Tables	+	+		
TalkExplorer (Verbert et al., 2013)					
SetFusion (Parra et al., 2014)					
TasteWeights (Bostandjiev et al., 2012)					
Moodplay (Andjelkovic et al., 2019)					
SmallWorld (Gretarsson et al., 2010)					
LinkedVis (Bostandjiev et al., 2013)					
(Schaffer et al., 2015)					
(Vlachos & Svonava, 2012)					
PARIS-ad (Jin et al., 2016)					
(Bakalov et al., 2013)					
(Kangasrääsiö, et al. 2015)					
System U (Badenes et al., 2014)					
UniWalk (Park et al., 2017)					
(Millecamp et al., 2019)					
RelExplorer (Tsai & Brusilovsky, 2017)					
Pharos (Zhao et al., 2011)					
(Zhang et al., 2014)					

4.1 What: Data Abstraction

As input for the visualizations, the surveyed tools focused on four different dataset types, namely tables, networks, sets, and geometry (see Figure 1). Tables are used in 12 tools to store different items/attributes, such as song attributes and user preferences for these attributes (Millecamp et al., 2019), artists and moods (Andjelkovic, Parra, & O'Donovan, 2019), publications, topics, co-authorships, and interests (Tsai & Brusilovsky, 2019), movies (Symeonidis, Nanopoulos, & Manolopoulos, 2009), Facebook profile attributes and personality traits (Jin, Seipp, Duval, & Verbert, 2016). Networks as a dataset type are mainly used in the tools providing recommendations based on social data. Examples of nodes/links in these networks include authors and co-authorships (Tsai & Brusilovsky, 2019), tags, recommender agents, and users (Verbert, Parra, Brusilovsky, & Duval, 2013), music, Wikipedia items, Facebook friends, and Twitter experts (Bostandjiev, O'Donovan, & Höllerer, 2013), movies (Vlachos & Svonava, 2012). Sets are used in many tools providing content-based explanation based on e.g. topics (Tsai & Brusilovsky, 2019), keywords (Tsai & Brusilovsky, 2017, 2019; Zhao et al., 2011), feature-opinion word pairs (Zhang et al., 2014). Only one tool uses the geometry dataset type to plot cities of affiliations on a world map, inspired by location-based explanation (Tsai & Brusilovsky, 2019).

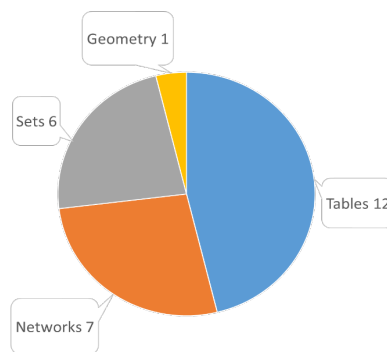


Figure 1: Dataset types

4.2 Why: Task Abstraction

As mentioned in Section 3.2, a task abstraction is expressed as an {action, target} pair. Ten tools out of the 21 surveyed tools are focusing on the {identify, paths} task (see Figure 2). These are mostly the tools which use networks as dataset type. For example, Relevance Tuner+ uses a path graph to visualize different possible connections between two conference attendees based on co-authorship information (Tsai & Brusilovsky, 2019). A layer-based interface connected via outgoing links is used by TasteWeights and LinkedVis (three layers) as well as SmallWorld (five layers) to visualize connections between the user profile and the recommended items (Bostandjiev, O'Donovan, & Höllerer, 2012; Bostandjiev et al., 2013; Gretarsson, O'Donovan, Bostandjiev, Hall, & Höllerer, 2010). PeerChooser incorporates a force-directed graph layout technique to visualize connections between an active user and different user-communities/peer-groups (O'Donovan, Smyth, Gretarsson, Bostandjiev, & Höllerer, 2008). Additionally, three of these tools focus on the {identify, topology} task to find clusters in the network based on user preferences (Gretarsson et al., 2010; O'Donovan et al., 2008; Verbert et al., 2013). In order to provide explanations based on similarity of items or attributes, another common task in the surveyed tools is {compare, similarity} between liked and

recommended music artists (Kouki et al., 2019), active user and recommended users as well as liked and recommended items (Park, Jeon, Kim, Ahn, & Kang, 2017), users' song preferences and recommended songs (Millecamp et al., 2019), searched keyword and feedback (Kangasrääsiö et al., 2015) or {summarize, similarity} between all artists (Kouki et al., 2019), artist moods (Andjelkovic et al., 2019), research papers (Parra, Brusilovsky, & Trattner, 2014), topics discussed in content-centric Web sites (Zhao et al., 2011). The {identify, features} and the {summarize, features} tasks are addressed by three tools each. For example, in (Zhang et al., 2014), the authors visualize the feature-opinion pairs of the recommended items as a tag cloud enabling users to identify the sentiment related to a specific pair of interest or get an overview of sentiments related to all the pairs. Similarly, in (Bakalov et al., 2013), the authors focus on both of these tasks to identify clusters and summarize interest degrees by visualizing items of interest categorized by different ontology classes on an IntrospectiveViews which is divided into zones and slices. Only two surveyed tools focus on the {identify, distribution} task. Tagsplanation shows the distribution of tags related to a specific movie in a bar chart (Vig et al., 2009) and System U presents the distribution of the personality traits of a user via a sunburst chart (Badenes et al., 2014).

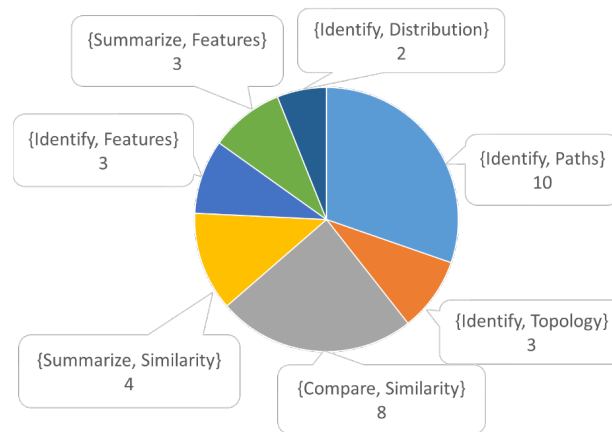


Figure 2: Task abstraction as {action, target} pairs

4.3 How: Idioms

This section discusses the visualization idioms used in the surveyed tools. Different charts, such as node-link diagrams, tag clouds, and bar charts are used to encode the explanatory visualizations and various interaction techniques are provided to enable users to interact with and control the visualizations.

4.3.1 Encode

To *encode* the explanatory visualizations, different charts are used by the tools that we have surveyed as shown in Figure 3. The most commonly used visualization idiom is the node-link diagram, which is used by 11 tools out of the 21 surveyed tools to indicate the relationships between items. For instance, the tool in (Vlachos & Svonava, 2012) represents movies as nodes and visualizes the connection between them based on their similarity. TasteWeights, LinkedVis, SmallWorld, and the tool in (Schaffer et al., 2015) aggregate related items representing nodes in the form of parallel layers and visually explain the resulted recommendations via connections (Bostandjiev et al., 2012, 2013; Gretarsson et al., 2010). PARIS-ad uses a simple node-link diagram consisting of four nodes to

explain the recommendation process based on the user profile (Jin et al., 2016). To summarize, the node-link diagram is an effective explanatory visualization for network-based datasets to enable efficient path identification and topology exploration.

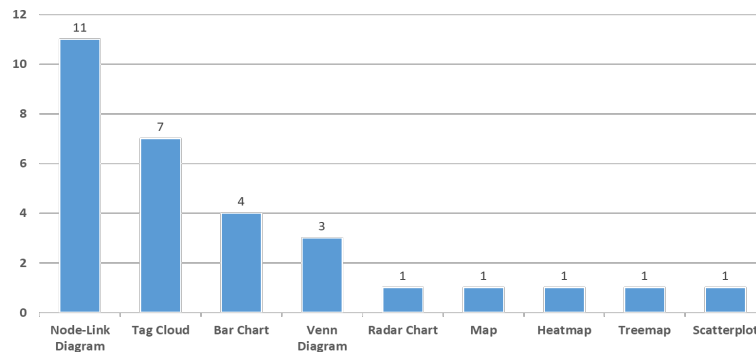


Figure 3: Chart types

Tag cloud is the second most used visualization idiom, which is used by seven surveyed tools to represent textual data. RelExplorer uses a tag cloud to assist co-attendees of an academic conference in understanding the level of content similarity between their publications and target scholars (Tsai & Brusilovsky, 2017). Similarly, the tool in (Zhang et al., 2014) provides the explanation of recommendations in the form of feature-opinion word pairs, which are further distinguished based on positive (green) and negative (blue) sentiments. Pharos uses a set of tag clouds, representing communities focusing on a similar topic, to summarize users' social behavior in content-centric Web sites, where each tag cloud visualizes the topics in green and people in blue (Zhao et al., 2011). In general, tag clouds are suitable for summarizing or comparing features similarity in set-based datasets.

Three surveyed tools use a venn diagram to show possible intersections between their datasets. SetFusion uses it to visualize recommended talks related to the publications in a conference which are common between three different sets containing publications similar to user's favorite, most bookmarked, and from frequently cited authors (Parra et al., 2014). Similar to SetFusion, HyPER uses a venn diagram to recommend intersecting music artists between sets of artists from a user profile, popular artists, and artists liked by people listening to similar artists as the user. Additionally, HyPER uses a tag cloud in each circle of the venn diagram to visualize keywords from a particular set (Kouki et al., 2019). Relevance Tuner+ also uses a venn diagram together with a tag cloud to explain similarity between the publications of a user and attendee of a conference (Tsai & Brusilovsky, 2019). To sum up, venn diagrams alone are good for comparing items and features, but together with tag clouds, they can provide effective visualizations to explain content-based datasets.

The third most used visualization idiom is the bar chart, which is used by four tools to compare similarity between user preferences and recommendations. For instance, to justify the recommended movie based on community tags, Tagsplanation uses bar charts to visualize each tag's relevance with the movie and the user's preference (Vig et al., 2009). UniWalk uses bar charts to provide explanation of recommended items based on similar users who liked those items and similar items liked by the user (Park et al., 2017). Similarly, the tool in (Millecamp et al., 2019) uses a grouped bar chart to show a comparison between recommended song's attributes and user's

preferences. Overall, bar chart is one of the most popular techniques to provide comparisons and similarities in table-based datasets.

The remaining five visualizations are used by one tool each. A radial heatmap called IntrospectiveViews is used in the tool proposed in (Bakalov et al., 2013) to visualize items based on the degree of interest and the type of item recommended. Most interesting items are displayed in the central hot zone while the least interesting items are presented in the cold zone on the edge of the radial heatmap. System U uses a radial treemap to visualize user's personality portraits including big five personality traits, fundamental needs, and basic human values (Badenes et al., 2014). The tool in (Millecamp et al., 2019) uses a scatterplot to show a correlation between two selected attributes of recommended songs. Finally, Relevance Tuner+ uses a radar chart to visualize research topics of a user compared to a conference attendee and a map to provide geo-distance, regions or countries information of similar conference attendee (Tsai & Brusilovsky, 2019).

4.3.2 Interact

The majority of the surveyed tools provide interactivity to help users interact with the explanatory visualizations (see Figure 4). In order to *manipulate* views, 16 out of 21 surveyed tools provide the ability to select an element to provide more details (Kouki et al., 2019; Tsai & Brusilovsky, 2019), highlight network paths (Bostandjiev et al., 2012, 2013; Gretarsson et al., 2010; Schaffer et al., 2015). Four tools allow users to change the view by moving a selected keyword to bring similar items closer (Kangasrääsiö et al., 2015), adding and removing network nodes to update recommendations and rearrange network (O'Donovan et al., 2008). Three tools implemented the zoom and pan functionality e.g. to view cities of affiliations on a world map (Tsai & Brusilovsky, 2019) or to more closely inspect areas of interest on mood space visualization (Andjelkovic et al., 2019). To support *reduce* interactions, 12 tools incorporated different filtering mechanisms to modify the visualization view using checkboxes (Andjelkovic et al., 2019; Bakalov et al., 2013; Verbert et al., 2013), sliders (Bakalov et al., 2013; Bostandjiev et al., 2012, 2013; Millecamp et al., 2019; Parra et al., 2014; Schaffer et al., 2015), buttons (Bostandjiev et al., 2012, 2013; O'Donovan et al., 2008; Schaffer et al., 2015), or dropdown menu (Bakalov et al., 2013; Millecamp et al., 2019). Eleven tools focused on *facet* interactions by providing juxtaposing and coordinating multiple views through selecting a topic to show recommended content and people (Zhao et al., 2011), a recommended user or item to show explanation (Park et al., 2017), or a region on a global view to get a detailed view (Bakalov et al., 2013).

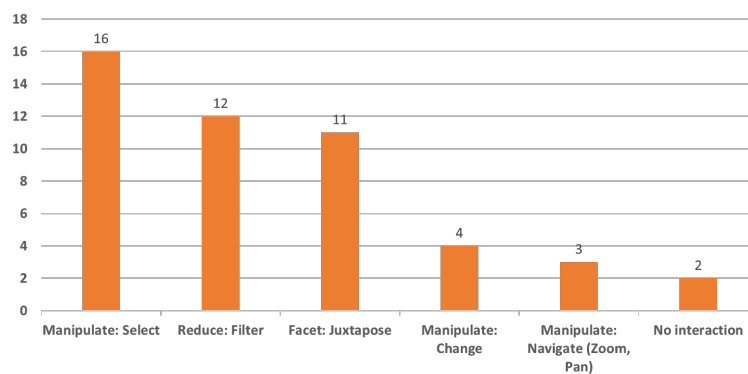


Figure 4: Interactivity types

5 DISCUSSION

Based on the results of our review, we present our insights and future lines of work in the field of visual explanations in recommender systems. In the design of explanatory visualizations, there are important design decisions to be made at the level of visual encoding and interaction with the explanation. These design decisions should be guided by design practices from the information visualization field (e.g. Munzner, 2014). In particular, the decision of what idiom should be used to encode the explanatory visualizations is driven by the data type as well as on the reason behind using the visualization. In other words, the data type and the visualization task impose what type of idiom should be used to visualize the explanation. For example, when the reason for using visualization is to compare similarity between items or users, the bar chart idiom presents a better choice when the data type is table, while a tag cloud / venn diagram is a more effective choice when the data type is set. When dealing with social and network data types, it is recommended to use the node-link diagram idiom for path identification and/or topology exploration tasks. As an important line of future work, there is a need to investigate how to adapt and apply information visualization design practices in visual explainable recommendation research, thus leading to more systematic research on explanatory visualizations in recommender systems.

Furthermore, providing interactivity with the explanatory visualizations is important to help users better understand the explanations. For example, juxtaposing and coordinating different views showing the user model, the recommendation algorithm, and the recommendation output can help users explore the relationship between their user model and the recommended items. Thus, there is a need to investigate in future works what kind of interaction should be supported to achieve effective explanations.

Another research question is how to relate explanatory visualization design with the aim (e.g. transparency, persuasiveness, scrutability) and type of explanation (e.g. collaborative filtering, content-based, social). An important direction for future work is to conduct evaluations aiming at answering the question “what kind of visual explanation is more effective for what kind of explanation aim and type?”.

6 CONCLUSION

In this paper, we highlighted the importance of explanations to increase transparency and trust in learning analytics. We focused on explainable recommendation as a possible solution to achieve explainable learning analytics (XLA). We systematically reviewed the research on explainable recommendation based on explanation display. Given the importance of visualization in the domain of learning analytics, we focused on visualization as a display style of explanation. Research on leveraging explainable recommendation in the domain of learning analytics is still in its initial stage, and there is much more to be explored in the future. We hope that with this survey, learning analytics researchers and practitioners will get a big picture of the visually explainable recommendation approaches as well as a better understanding on how to adapt and apply this research to develop explainable, transparent, and trustworthy educational recommender systems.

REFERENCES

- Andjelkovic, I., Parra, D., & O'Donovan, J. (2019). Moodplay: Interactive music recommendation based on Artists' mood similarity. *International Journal of Human-Computer Studies*, 121, 142–159. <https://doi.org/https://doi.org/10.1016/j.ijhcs.2018.04.004>
- Badenes, H., Bengualid, M. N., Chen, J., Gou, L., Haber, E., Mahmud, J., ... Zhou, M. X. (2014). System U: Automatically Deriving Personality Traits from Social Media for People Recommendation. *Proceedings of the 8th ACM Conference on Recommender Systems*, 373–374. <https://doi.org/10.1145/2645710.2645719>
- Bakalov, F., Meurs, M.-J., König-Ries, B., Sateli, B., Witte, R., Butler, G., & Tsang, A. (2013). An Approach to Controlling User Models and Personalization Effects in Recommender Systems. *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, 49–56. <https://doi.org/10.1145/2449396.2449405>
- Bostandjiev, S., O'Donovan, J., & Höllerer, T. (2012). TasteWeights: A Visual Interactive Hybrid Recommender System. *Proceedings of the Sixth ACM Conference on Recommender Systems*, 35–42. <https://doi.org/10.1145/2365952.2365964>
- Bostandjiev, S., O'Donovan, J., & Höllerer, T. (2013). LinkedVis: Exploring Social and Semantic Career Recommendations. *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, 107–116. <https://doi.org/10.1145/2449396.2449412>
- Drachsler, H., & Greller, W. (2016). Privacy and Analytics: It's a DELICATE Issue a Checklist for Trusted Learning Analytics. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 89–98. <https://doi.org/10.1145/2883851.2883893>
- Friedrich, G., & Zanker, M. (2011). A taxonomy for generating explanations in recommender systems. *AI Magazine*. <https://doi.org/10.1609/aimag.v32i3.2365>
- Gedikli, F., Jannach, D., & Ge, M. (2014). How should I explain? A comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies*, 72(4), 367–382. <https://doi.org/https://doi.org/10.1016/j.ijhcs.2013.12.007>
- Gretarsson, B., O'Donovan, J., Bostandjiev, S., Hall, C., & Höllerer, T. (2010). SmallWorlds: Visualizing Social Recommendations. *Computer Graphics Forum*, 29(3), 833–842. <https://doi.org/10.1111/j.1467-8659.2009.01679.x>
- He, C., Parra, D., & Verbert, K. (2016). Interactive recommender systems: A survey of the state of the art and future research challenges and opportunities. *Expert Systems with Applications*, 56, 9–27. <https://doi.org/https://doi.org/10.1016/j.eswa.2016.02.013>
- Jin, Y., Seipp, K., Duval, E., & Verbert, K. (2016). Go With the Flow: Effects of Transparency and User Control on Targeted Advertising Using Flow Charts. *Proceedings of the International Working Conference on Advanced Visual Interfaces*, 68–75. <https://doi.org/10.1145/2909132.2909269>
- Kangasrääsio, A., Glowacka, D., & Kaski, S. (2015). Improving Controllability and Predictability of Interactive Recommendation Interfaces for Exploratory Search. *Proceedings of the 20th International Conference on Intelligent User Interfaces*, 247–251. <https://doi.org/10.1145/2678025.2701371>
- Knijnenburg, B. P., Willemsen, M. C., Gantner, Z., Soncu, H., & Newell, C. (2012). Explaining the User

- Experience of Recommender Systems. *User Modeling and User-Adapted Interaction*, 22(4–5), 441–504. <https://doi.org/10.1007/s11257-011-9118-4>
- Konstan, J. A., & Riedl, J. (2012). Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction*, 22(1), 101–123. <https://doi.org/10.1007/s11257-011-9112-x>
- Kouki, P., Schaffer, J., Pujara, J., O'Donovan, J., & Getoor, L. (2019). Personalized Explanations for Hybrid Recommender Systems. *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 379–390. <https://doi.org/10.1145/3301275.3302306>
- Millecamp, M., Htun, N. N., Conati, C., & Verbert, K. (2019). To Explain or Not to Explain: The Effects of Personal Characteristics when Explaining Music Recommendations. *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 397–407. <https://doi.org/10.1145/3301275.3302313>
- Munzner, T. (2014). *Visualization Analysis and Design*. CRC Press.
- Nunes, I., & Jannach, D. (2017). A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction*, 27(3), 393–444. <https://doi.org/10.1007/s11257-017-9195-0>
- O'Donovan, J., Smyth, B., Gretarsson, B., Bostandjiev, S., & Höllerer, T. (2008). PeerChooser: Visual Interactive Recommendation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1085–1088. <https://doi.org/10.1145/1357054.1357222>
- Papadimitriou, A., Symeonidis, P., & Manolopoulos, Y. (2012). A generalized taxonomy of explanations styles for traditional and social recommender systems. *Data Mining and Knowledge Discovery*, 24(3), 555–583. <https://doi.org/10.1007/s10618-011-0215-0>
- Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450. <https://doi.org/10.1111/bjet.12152>
- Park, H., Jeon, H., Kim, J., Ahn, B., & Kang, U. (2017). Uniwalk: Explainable and accurate recommendation for rating and network data. *ArXiv E-Prints*.
- Parra, D., Brusilovsky, P., & Trattner, C. (2014). See What You Want to See: Visual User-driven Approach for Hybrid Recommendation. *Proceedings of the 19th International Conference on Intelligent User Interfaces*, 235–240. <https://doi.org/10.1145/2557500.2557542>
- Pu, P., Chen, L., & Hu, R. (2011). A User-centric Evaluation Framework for Recommender Systems. *Proceedings of the Fifth ACM Conference on Recommender Systems*, 157–164. <https://doi.org/10.1145/2043932.2043962>
- Pu, P., Chen, L., & Hu, R. (2012). Evaluating recommender systems from the user's perspective: survey of the state of the art. *User Modeling and User-Adapted Interaction*, 22(4), 317–355. <https://doi.org/10.1007/s11257-011-9115-7>
- Schaffer, J., Hollerer, T., & O'Donovan, J. (2015). Hypothetical recommendation: A study of interactive profile manipulation behavior for recommender systems. *The Twenty-Eighth International Flairs Conference*.
- Symeonidis, P., Nanopoulos, A., & Manolopoulos, Y. (2009). MoviExplain: a recommender system

- with explanations. *Proceedings of the Third ACM Conference on Recommender Systems*, 317–320.
- Tintarev, N., & Masthoff, J. (2015). Explaining Recommendations: Design and Evaluation. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook* (pp. 353–382). Boston, MA: Springer US.
- Tsai, C.-H., & Brusilovsky, P. (2017). Providing Control and Transparency in a Social Recommender System for Academic Conferences. *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, 313–317. <https://doi.org/10.1145/3079628.3079701>
- Tsai, C.-H., & Brusilovsky, P. (2019). Explaining Recommendations in an Interactive Hybrid Social Recommender. *Proceedings of the 24th International Conference on Intelligent User Interfaces*, 391–396. <https://doi.org/10.1145/3301275.3302318>
- Verbert, K., Manouselis, N., Ochoa, X., Wolpers, M., Drachsler, H., Bosnic, I., & Duval, E. (2012). Context-Aware Recommender Systems for Learning: A Survey and Future Challenges. *IEEE Transactions on Learning Technologies*, 5(4), 318–335. <https://doi.org/10.1109/TLT.2012.11>
- Verbert, K., Parra, D., Brusilovsky, P., & Duval, E. (2013). Visualizing Recommendations to Support Exploration, Transparency and Controllability. *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, 351–362. <https://doi.org/10.1145/2449396.2449442>
- Vig, J., Sen, S., & Riedl, J. (2009). Tagsplanations: Explaining Recommendations Using Tags. *Proceedings of the 14th International Conference on Intelligent User Interfaces*, 47–56. <https://doi.org/10.1145/1502650.1502661>
- Vlachos, M., & Svonava, D. (2012). Graph embeddings for movie visualization and recommendation. *First International Workshop on Recommendation Technologies for Lifestyle Change (LIFESTYLE 2012)*, 56.
- Zhang, Y., & Chen, X. (2018). Explainable Recommendation: {A} Survey and New Perspectives. *CoRR*, abs/1804.1. Retrieved from <http://arxiv.org/abs/1804.11192>
- Zhang, Y., Lai, G., Zhang, M., Zhang, Y., Liu, Y., & Ma, S. (2014). Explicit Factor Models for Explainable Recommendation Based on Phrase-level Sentiment Analysis. *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, 83–92. <https://doi.org/10.1145/2600428.2609579>
- Zhao, S., Zhou, M. X., Zhang, X., Yuan, Q., Zheng, W., & Fu, R. (2011). Who is Doing What and When: Social Map-Based Recommendation for Content-Centric Social Web Sites. *ACM Trans. Intell. Syst. Technol.*, 3(1), 5:1--5:23. <https://doi.org/10.1145/2036264.2036269>

Exploring the Design Space for Explainable Course Recommendation Systems in University Environments

Boxuan Ma, Min Lu, Yuta Taniguchi, Shin'ichi Konomi

Kyushu University, Japan

ma.boxuan.611@s.kyushu-u.ac.jp

ABSTRACT: A course recommendation system can assist students in finding suitable courses through a personalized approach. Showing why the course is recommended serves as a bridge between the recommender system and student, could increase student's trust in the system and persuade the student to accept the course. This paper presents a study of the factors that influence students' course selection in universities so as to better understand student perceptions, attitudes, and needs and leverage data-driven approaches for recommending and explaining the recommendations for complex and interactive university environments.

Keywords: Learning Analytics; Explainability; Recommender Systems

1 INTRODUCTION

Due to the increasing number of students and the rise of MOOCs, course recommender systems have become a well-researched area [Polyzou & Karypis, 2019; Jiang, Pardos & Wei, 2019]. Course recommendation is considered a useful tool in the education field for helping students who have no sufficient experience to choose courses that they need as well as reducing time to explore courses that they will take. These recommender systems are supported by the data, explicitly through, but also implicitly through for instance learning activity behavior. Current recommender systems often behave like a "black box" (e.g., those using deep neural networks [Pardos, Fan & Jiang, 2019]): i.e., recommendations are presented to the users, but the rationale for selecting recommendations is often not explained to end-users [Parra & Brusilovsky, 2015]. Several researchers have shown that explanations and user control are needed to support the interpretation of the data and decision-making [Kulesza et al., 2015].

In this context, this paper presents the results of our interviews and surveys to untangle the complex factors that are of concern to university students in order to inform the design of alternative course recommender systems that may consider the versatile nature of reasons involved in course selection.

2 RELATED WORK

2.1 Course selection

Some work has been done on analyzing the college students' course selection. [Kinnunen & Malmi, 2006] conducted a study on the reasons for students' quitting the CS1 course at Helsinki University of Technology. It was discovered that one of the most important factors

of quitting the CS1 course is the perceived difficulty of the course. This study suggests that it is a suitable strategy to recommend courses that students will obtain relatively higher course achievement. In [Tallón et al., 2014], a survey was conducted to analyze why students choose one elective course. However, it is limited to only the case of teratology. Kardan et al. [2013] conducted a study on the factors influencing online course selection of college students in the context of e-learning using a neural network. However, it is limited to the e-learning which may significantly differ from face-to-face learning in the traditional college education.

Additionally, there is still a lack of study on the factors that influence students' course selection in university and how the course selection would impact the students' educational achievement.

2.2 Course recommendation system

Recommendation systems have been broadly applied within the context of student learning. Our review of relevant literature shows that many of the recent works on course recommendation environments focus on online learning platforms such as MOOCs [Jing & Tang, 2017]. Other studies on course recommendation use datasets collected in physical university environments, however, they rely on recommendation approaches that are similar to the ones used in recommending MOOC courses without fully considering the versatile nature of the reasons involved in course selection in physically-based university environments [Jiang & Wei, 2019]. This amounts to a collaborative recommendation of the nature of "most people like you did X next." When it comes to students' diverse intentions in selecting courses, a student's goal may not align with what most people have done. Although a few existing works consider the characteristics of university environments, they tend to make simplistic assumptions about learners and their contexts, thereby merely recommending the whole sequence of courses that satisfy the degree requirements [Parameswaran, Venetis & Garcia-Molina, 2011], or predicting the performance of students and give recommendations based on predicted results [Elbadrawy & Karypis, 2016].

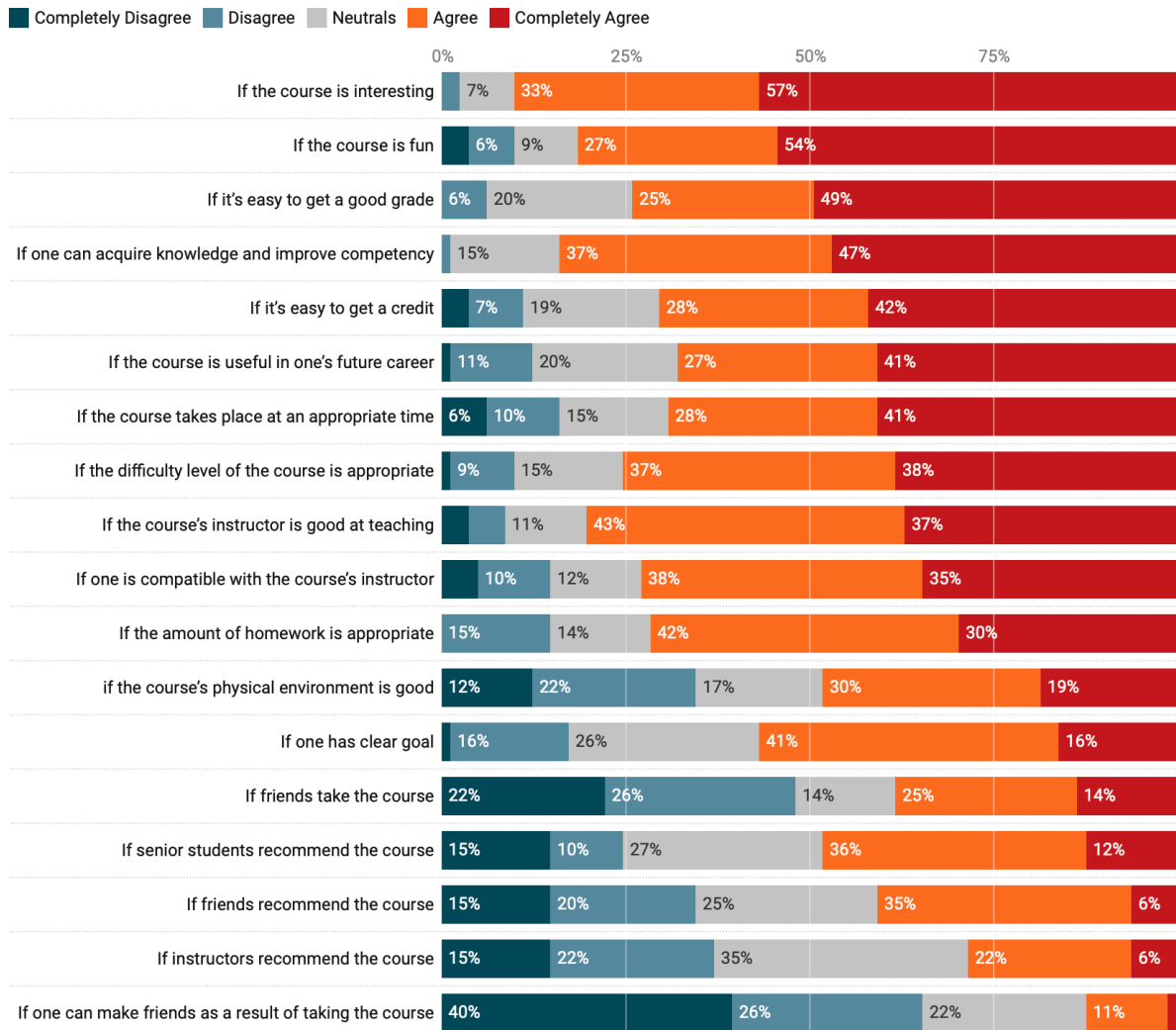
In this paper, we elaborate on the explainability of course recommendation and explore how to support the design space for explainable course recommendation systems in university environments.

3 PRELIMINARY STUDY

Intuitively, the reasons behind course selection are manifold. Likewise, students who are enrolled in the same course may have completely different orientations. To get a better understanding of course selection behavior, we conducted a preliminary study. We have employed a two-phased approach to analyze the factors that need to be considered to recommend courses in university environments. First, we conducted a qualitative interview with 10 students (N=10). In the interviews, we asked about their opinions on the course selection and the reasons behind their decisions to choose courses. This procedure has revealed various reasons for course selection and allowed us to extract potential factors for effective course recommendations in university environments. Second, the extracted factors have been used to design 18 survey questions that ask respondents to rate the importance of each factor in selecting courses.

Table 1: Questionnaire regarding course selection behavior¹.

Item	Description	Item	Description
Q1	If it's easy to get a credit	Q10	If the course's instructor is good at teaching
Q2	If it's easy to get a good grade	Q11	If one is compatible with the course's instructor
Q3	If the difficulty level of the course is appropriate	Q12	If the course is fun
Q4	If one can acquire knowledge and improve competency	Q13	If the course takes place at an appropriate time
Q5	If the course is useful in one's future career	Q14	If friends take the course
Q6	If the course is interesting	Q15	If one can make friends as a result of taking the course
Q7	If friends recommend the course	Q16	If the amount of homework is appropriate
Q8	If senior students recommend the course	Q17	if the course's physical environment is good ²
Q9	If instructors recommend the course	Q18	If one has clear goal

**Figure 1: Survey results.**¹ These are the English translation of the description, which is originally in Japanese.² Such as temperature, humidity, WIFI connectivity.

This study took place at Kyushu University and participants are all students. Kyushu University is a national university in Japan and it offers a wide variety of degree programs. The questionnaire was sent to all (N=336) students from two information science courses at Kyushu University and 24.1% (N=81) of students responded to the survey (53 men, 28 women; 78 freshmen, 1 sophomore, 1 junior, 1 senior). A 5-Likert scale (1-completely disagree, 5-completely agree) questionnaire regarding student requirement in terms of course selection behavior (Nstudent=81) with the questions presented in Table 1.

3.1 General Motivation

Course recommendation for higher education is “messy and unorganized” [Babad & Tayeb, 2003] as it depends on many factors that students need to concern. Table 1 indicates that many factors affect students’ decision to choose courses and Figure 1 shows survey results. The overall most important factor was students’ interest (Q6). Furthermore, the factors relevant to students’ career goals (Q5) and grades (Q2) are perceived to be important in course selection. From the above results, one safe conclusion can be drawn that, there are complex constraints and contexts that have to be considered together and students have to balance all those factors above, made more difficult by the multiple objectives that students want to maximize and risks they want to hedge against. For example, choosing challenging courses of value while maintaining a high GPA.

3.2 Learning Goal and Career Plan

University environments are inherently different from MOOC environments. For example, MOOC users may have clearer learning goals than university freshmen who may still be exploring different possibilities in relation to their careers and learning. Also, physical and social university environments provide students with a plethora of opportunities to explore, discover and develop intellectual interests and meaningful goals.

Answer to Q18 revealed that 17.2% of first-year students are either very unclear or unclear about their learning goals, and 26% of first-year students are neutrals and don't know about this question, the choice of courses for those students is aimless. Also, student interest and goal can change as they explore and discover something meaningful on and off campus. This kind of students has different criteria for course selection than the students who have a clear learning goal. Figure 2 shows the sample of our survey results comparing the students with clear learning goals (the orange bar) and the students without clear learning goals (the blue bars). Students with clear learning goals considered that the usefulness and relevance to their future goals (Q4, Q5) are important in course selection. Also, they would like to ask their professors for advice (Q9). It may be because advice from professors who tend to be local experts with deep knowledge within their subject area is useful. In contrast, students without clear goals considered ease of getting credits (Q1) as important. Also, they prefer to ask suggestions from their friends (Q7). Those students who have no clear goals might be more inclined to choose easy courses that do not require too much effort. They may also exploit social means of obtaining recommendations more than the students with clear goals.

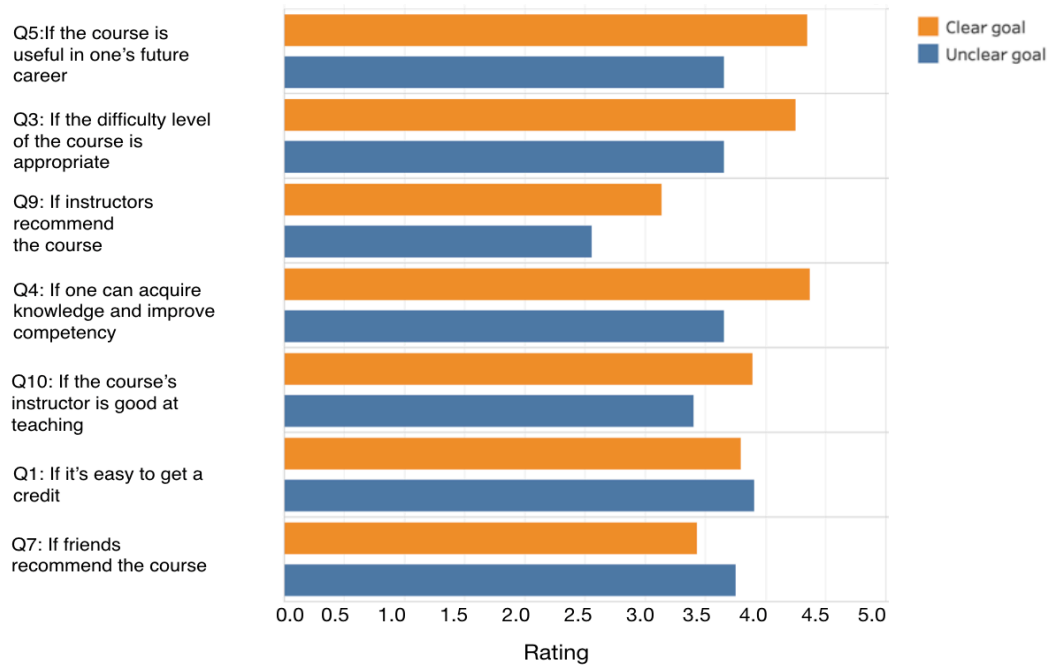


Figure 2: Comparison results between students with clear learning goals and students without clear learning goals.

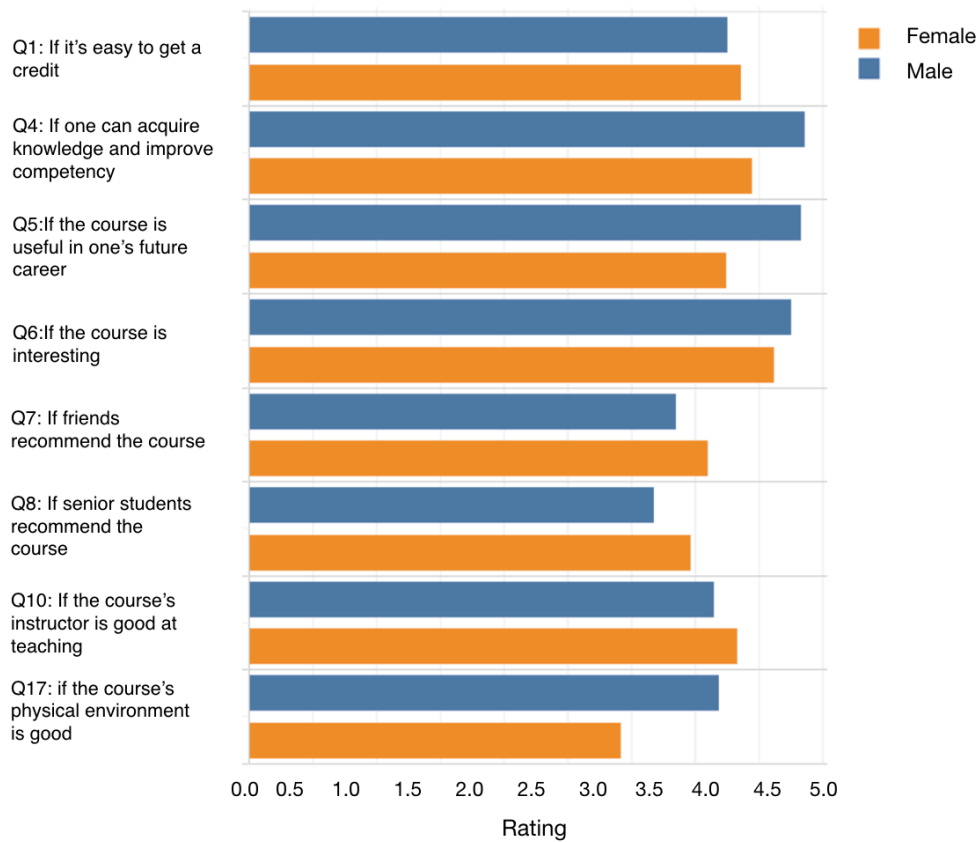


Figure 3: Comparison results between female students and male students.

3.3 Social Aspect

Potts et al. [2018] conclude that the risk of social isolation is a problem in the learning process especially for first-year students at university, who have difficulty navigating their new academic and environment. In fact, the social factor also plays a part in the course selection process. For example, some students prefer to enroll in a course with their friends or classmates together. Tinto [1997] concludes that participation in a collaborative learning group enables students to develop a network of support. This community of classroom-based peers encourages student's attendance and class participation.

Our results show that 39% of first-year students considered the social factor as an important thing in their course selection process. The results in Figure 3 also show some interesting gender differences regarding the social aspect. Female students seem to ask suggestions from their friends and senior students more than male students (Q7, Q8). Also, female students consider factors related to the instructors of courses as important more than male students (Q10). The above results indicate that the classmates or friends based social link could be important information in course recommendation.

3.4 Student Preference

We also analyzed the survey data by employing the k-means clustering algorithm to identify different types of students in terms of course selection motivations. Table 2 shows the clustering results. It can be seen that students of cluster 1 consider high grade as the most important factor (Q2, M=5.74) and they are inclined to choose courses that do not require too much effort or difficulty (Q1, M=4.65; Q3, M=4.30). Students of cluster 2 prefer to choose courses to pave the way for their future career (Q4, M=4.58; Q5, M=4.47), but they also want to make a trade-off between high GPA (Q2, M=4.13) and their interest (Q6, M=4.34). Students of cluster 3 seem to be challengers, as they may even take difficult courses (Q3, M=3.57) if they are interested in them (Q6, M=4.27) or think the course is helpful for their future career (Q4, M=4; Q5, M=4.22).

Table 2: Sample clustering results.

Item	Q1	Q2	Q3	Q4	Q5	Q6
Cluster1	4.65	4.74	4.30	3.57	3.22	3.43
Cluster2	3.72	4.13	4.25	4.58	4.47	4.34
Cluster3	3.09	3.70	3.57	4.00	4.22	4.27

4 DISCUSSION AND CONCLUSION

Although conducted on a relatively small-scale, our preliminary study has revealed the complexity and variety of factors involved in students' decision to choose courses in the university environments. Existing course recommendation approaches without considering such factors may not fit different perceptions of students, and therefore could not provide convincing explanations that are needed to support the interpretation of the data and decision-making. Our survey results show that different students may have completely different orientations based on their own reasons, which serves as different criteria for

course selection and those should be considered in course recommendation systems in physically-based learning environments such as universities. This suggests that recommendations and explanations that are aimed only at one or a few factors are likely not enough to help the students.

Based on our results, it is important for course recommendation approaches to: (a) take different factors into account when training models. e.g., social factor and physical constraints in university environments; (b) consider the courses are well-aligned with their interest and learning goal; and (c) account for the student's expected grades in each course that they recommend. Accordingly, physically-based course recommendation systems should provide relevant information to explain the recommendations: (a) course descriptions and structures that make students have a better understanding about the course (b) personal preference of students (e.g., their interest, major, the courses they have already taken and their grades) and (c) interactive representations that help students make sense of recommended courses from multiple perspectives.

Our future research will concentrate on two issues. First, we shall carry out a more specific analysis of large-scale data on students' selection decisions and a more detailed analysis of the selection process. In addition, we shall link the results to the actual use in building the explainable course recommender system.

REFERENCES

- Babad, E., & Tayeb, A. (2003). Experimental analysis of students' course selection. *British Journal of Educational Psychology*, 73(3), 373-393.
- Elbadrawy, A., & Karypis, G. (2016). Domain-aware grade prediction and top-n course recommendation. In *Proceedings of the 10th ACM Conference on Recommender Systems* (pp. 183-190). ACM.
- Jiang, W., Pardos, Z. A., & Wei, Q. (2019). Goal-based course recommendation. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 36-45). ACM.
- Jing, X., & Tang, J. (2017). Guess you like: course recommendation in MOOCs. In *Proceedings of the International Conference on Web Intelligence* (pp. 783-789). ACM.
- Kardan, A. A., Sadeghi, H., Ghidary, S. S., & Sani, M. R. F. (2013). Prediction of student course selection in online higher education institutes using neural network. *Computers & Education*, 65, 1-11.
- Kinnunen, P., & Malmi, L. (2006). Why students drop out CS1 course?. In *Proceedings of the second international workshop on Computing education research* (pp. 97-108). ACM.
- Kulesza, T., Burnett, M., Wong, W. K., & Stumpf, S. (2015, March). Principles of explanatory debugging to personalize interactive machine learning. In *Proceedings of the 20th international conference on intelligent user interfaces* (pp. 126-137). ACM.
- Parameswaran, A., Venetis, P., & Garcia-Molina, H. (2011). Recommendation systems with complex constraints: A course recommendation perspective. *ACM Transactions on Information Systems (TOIS)*, 29(4), 20.

- Pardos, Z. A., Fan, Z., & Jiang, W. (2019). Connectionist recommendation in the wild: on the utility and scrutability of neural networks for personalized course guidance. *User Modeling and User-Adapted Interaction*, 29(2), 487-525.
- Parra, D., & Brusilovsky, P. (2015). User-controllable personalization: A case study with SetFusion. *International Journal of Human-Computer Studies*, 78, 43-67.
- Polyzou, A., Athanasios, N., & Karypis, G. (2019). Scholars Walk: A Markov Chain Framework for Course Recommendation.
- Potts, B. A., Khosravi, H., Reidsema, C., Bakharia, A., Belonogoff, M., & Fleming, M. (2018). Reciprocal peer recommendation for learning purposes. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 226-235). ACM.
- Tallón, V., Carvalho, P., Vinyals, H., & Manzanares, M. (2014). Why do they choose an optional course? The rare case of teratology. *European Journal of Dental Education*, 18(2).
- Tinto, V. (1997). Classrooms as communities: Exploring the educational character of student persistence. *The Journal of higher education*, 68(6), 599-623.

Explainable Learning Analytics: challenges and opportunities

Tinne De Laet , Martijn Millecamp, Tom Broos, Robin De Croon, Katrien Verbert

KU Leuven, Belgium

{tinne.delaet,martijn.millecamp,tom.broos,robin.decroon,katrien.verbert}@kuleuven.be

Raphael Duorado

Universidade Federal de Pernambuco, Brasil

rasd2@cin.ufpe.br

ABSTRACT: In the last decade, we are witnessing a widespread adoption of artificial intelligence in a wide range of application domains. Learning analytics is no exception. Artificial Intelligence (AI) techniques and Machine Learning (ML) in particular are used to generate automatic predictions and recommendations regarding learning and teaching. A key challenge in the actual use and adoption of AI and ML is that they often operate as a ‘black box’, hereby impeding understanding and trust. The domain of Explainable Artificial Intelligence aims at enhancing the transparency of AI techniques and therefore also holds substantial promise for the Learning Analytics domain. This paper supports the shaping of the research line of Explainable Learning Analytics (XLA), by exploring the challenges and opportunities related to the data, stakeholders, communication, evaluation, and implementation & adoption of XLA. In particular, this paper reports on the outcomes of a 3-hour workshop with 44 international participants in which these challenges and opportunities were collaboratively identified. The obtained challenges and opportunities will form the basis for a deeper exploration, involving a wide range of stakeholders, of the promises of the XLA-field and the required points of focus for the next 10 years.

Keywords: learning analytics, explainable learning analytics, explainable artificial intelligence, recommender systems, visual analytics



Figure 1: Example of workshop outcome: opportunities and challenges of explainable learning analytics regarding the stakeholder dimension. The stickers are the result of an up- and downvoting procedure (green = upvoted challenge, gold = upvoted opportunity, yellow = veto).

1 INTRODUCTION

The domain of Learning Analytics (LA) has been established at the background of the never-ending need for support of learners and teachers and under impulse of the growth of learning data, the development of algorithms and AI, and the research within the learning sciences. LA is about “collecting traces that learners leave behind and using those traces to improve learning” (Duval, 2012). Obtaining actual improvement in learning and teaching does not come easy however. Verbert et al. (2013) introduced a LA process model consisting of four levels: awareness (level 1), (self-) reflection (level 2), sensemaking (level 3), and impact (level 4). This model shows that impact in the form of behavioral changes, new meanings, and improvements can only be obtained after awareness of the data, (self-)reflection, and sense-making have been obtained. Machine Learning (ML) and Artificial Intelligence (AI) have provided plenty of algorithms that can support the translation from data to awareness, self-reflection, and sense-making. Visualization techniques in general and Learning Analytics Dashboards (LAD) in particular (Verbert et al., 2014) have been shown to be able to provide a visual means of communication of the data and the outcomes of AI algorithms to stakeholders. While the rapidly maturing LA domain has proven to be an interesting domain of application for AI, ML, and visual analytics, it is creating higher expectations regarding the algorithmic predictions and recommendations generated by AI, ML, and visual analytics. These are expected to be interpretable for and explainable to the involved stakeholders and should be able to be translated to actionable recommendations.

To be **interpretable and explainable**, the outcomes of the data analysis, visualization, and/or ML and AI algorithms often have to be tailored to the particular stakeholders and end-users. While advanced visualization and/or ML techniques might create accurate and trustworthy insights and recommendations, this not automatically leads to the users trusting their outputs. Opening the black-box of LA to the user, in a user-tailored fashion, is an important step towards obtaining interpretable insights and explainable recommendations. The use of new approaches to obtain transparency, trustworthiness, persuasiveness, and effectiveness support this evolution.

A second challenge for LA, after interpretability and explainability, is to translate predictions and recommendations into feasible ‘actions’. This is also referred to as **actionability**. To highlight the challenge of this actionability, let’s consider the following example. Educational data mining techniques may discover that male students on average are more likely to fail in higher education. While such information can be interesting for researchers and policy makers, it does not provide a directly actionable recommendation towards an individual (male) student on how to improve his learning or study success. If actionable insights and recommendations can be created within LA and if they can be tailored to the involved stakeholders, they will have the potential to create impact (Verbert et al., 2013). User-centered design involving the stakeholders and the integration of LA into actual educational practices and in the pedagogy underlying these educational practices will support the actionability of the insights and recommendations. For example; if instructors collaborate with ML researchers when integrating LA in the form of automatic resource recommendation in their course design, they can help to understand and interpret the automatic recommendation of resources to students in the context of a particular class.

The domain of explainable Artificial Intelligence (XAI) has been developing and growing fast in the last years. AI is a part of a new generation of AI technologies called the third wave AI including,

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

among other ambitious goals, the development of algorithms that can explain themselves (Adadi & Berrada, 2018). The XAI research field aims at improving trust and transparency of AI-based systems, which can both concern automatic predictions and recommendations. AI algorithms often suffer from the so-called 'black-box' phenomenon, indicating that it is hard for users, including domain experts, to get insights in the internal mechanisms underlying the algorithms and the outcomes produced by these algorithms. This is also referred to as algorithm opacity (Adadi & Berrada, 2018). The problem of opacity has been growing together with the development of novel ML algorithms, such as deep learning and random forests, which itself was supported by the rapidly growing computational power. Algorithm opacity can however impede trust in predictions and recommendations provided by these ML algorithms and AI techniques, preventing their actual adoption and deployment in real-world scenarios.

The evolution towards actionable insights and explainable recommendations is urgent, as recent data protection and privacy regulations like the EU General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) stipulate that transparency is a fundamental right. In this light, the use of 'black box' approaches towards end-users becomes more and more challenging as the algorithmic approaches themselves lack transparency for LA end-users. Even more, GDPR states that users have the right to withdraw themselves from automatic decision making and profiling. Human mediation in automatic decision making and profiling is a promising approach to accommodate for the ethical use. Human mediation can however only reach its full potential if the algorithmic outputs are interpretable and explainable by the human mediator.

The XAI field by itself is rapidly maturing as shown in the XAI survey of Adadi and Berrada (2018) and by even more recent contributions focusing on the trends within XAI (Abdul et al. , 2018), on the sub-domain of explainable recommendations (Y. Zhang & Chen, 2018; Ouyang, Lawlor, Costa, & Dolog, 2018), the evaluation of XAI (Mohseni, Zarei, & Ragan, 2018), and visual interpretability (Spinner, Schlegel, Schäfer, & El-Assady, 2019; Zhao, Wu, Lee, & Cui, 2019; Q.-s. Zhang & Zhu, 2018; Choo & Liu, 2018).

The survey of Adadi and Berrada (2018) also recognizes the potential for XAI in different application domains: transportation, healthcare/medical (Vellido, 2019; Kwon et al., 2019), legal, finance, and military. The number of application domains is still growing fast as shown by recent research dedicated to e.g. robotic agents (Anjomshoei, Najjar, Calvaresi, & Frmling, 2019). While attention for XAI is also growing within the domain of LA, it still remains to be determined what the main research directions should be, to what level general XAI findings can be applied to the LA domain, and to what level specific developments have to be made.

The goal of this paper is to contribute to the creation and shaping of the exciting and promising domain of Explainable Learning Analytics (XLA), focusing on the application domain-specific developments of XAI within LA. In particular, this paper aims at contributing to the discovery of the main opportunities and challenges of XLA. To this end, this paper reports on a workshop involving more than 40 international stakeholders to identify the main challenges and opportunities of XLA.

2 METHODOLOGY

Stakeholder input regarding the challenges and opportunities of XLA was collected during a 3-hour workshop at the 2019 European Conference on Technology Enhanced Learning (EC-TEL 2019), in Delft, the Netherlands. Beforehand, the involved researchers identified five themes of focus for the workshop: data, stakeholders, communication, evaluation, and implementation & adoption. Table 1 provides an overview of the themes and how they were presented during the workshop.

Table 1: The five different themes of focus of explainable learning analytics and the teaser questions that were provided to the participants.

Theme	Explanation verbally provided to participants
Data	What data can be useful to be explained? What data about the user can be used to generate a prediction or a recommendation? Is the data readily available?
Stakeholders	Who are the stakeholders of explainable LA? Why would they use LA that needs explanations? In which situation would they need explanations? When would they see/use these explanations?
Communication	How would you communicate explanations to the user? (visual, text, mix, audio, ...). How would you adapt the explanation? Based on personal/situational characteristics? Is it ethical to adapt interfaces? Why would you trust or not trust a system? Do explanations help?
Evaluation	How would you evaluate explanations? Qualitative? Quantitative? What end goal would you evaluate?
Implementation & adoption	What steps are needed to implement XLA? Where do you see the domain of XLA in 10 years? What is needed to reach that? What can inhibit XLA? What can stimulate XLA? For what purpose would you use XLA?

These themes are inspired by the six critical dimensions of LA of Greller & Drachsler (2012) (data, stakeholders, instruments, internal limitations, external constraints) supplemented with and made more concrete by experience of the researchers themselves in the implementation at institutional scale of student dashboards (Broos et al., 2020). The data and stakeholder theme were directly borrowed from the six critical dimensions. Communication is one specific aspect of the critical dimension ‘instruments’, focusing on the communication of algorithmic predictions and recommendations in the context of the workshop. The theme of implementation & adoption touches on the internal limitations and external constraints of the six critical dimensions. However, we decided, based on our experience with deploying learning dashboards at institutional scale to focus specifically on implementation & adoption. Finally, the theme of evaluation concerns both the evaluation methodology (Instruments dimension), but also what final objective (Objectives dimension) to evaluate. These themes were used to structure the conversation and we do not claim that these five themes entail all possible viewpoints of XLA, nor that they are the only way to structure them.

During the workshop the following protocol was used: (1) Welcome and ice-breaker activity (10 minutes); (2) Introduction regarding LA, explainability and interpretability of predictions and recommendations (15 minutes), (3) Idea generation round in small groups (60 minutes, Figure 2), (4) Synthesis round where all input per theme was collected and prioritized using grouping of input and

up and downvoting, (30 minutes, Figure 1), (5) Plenary discussion to finalize the identified challenges and opportunities (65 minutes).



Figure 2: Picture of the idea generation round during the workshop at the EC-TEL 2019 conference

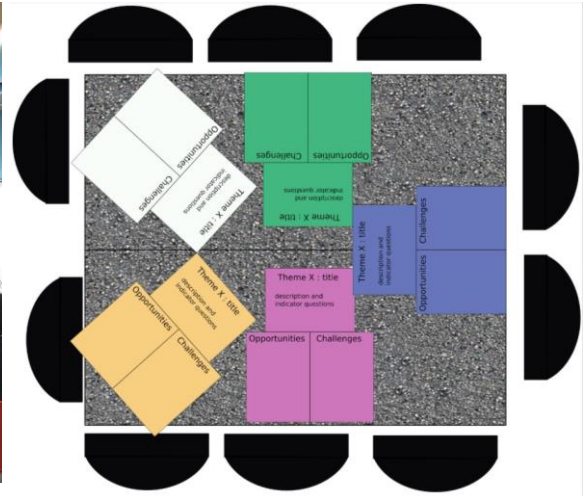


Figure 3: Table lay-out for the idea generation round. At each table the five themes are discussed by pairs of participants.

In the *introduction* a plenary presentation was provided with the general background of explainable AI, LA, and some examples of XLA. The presentation also focused on the concepts of predictions, recommendations, explainability, and interpretability and why these are important. This introduction ensured that each participant had a basic understanding of the field of XAI and LA. Additionally, everyone was made aware of the protocol used in the workshop.

The *idea generation round* was organized such that pairs of participants discussed around each of the five themes. Tables were prepared to support this, as illustrated in Figure 3. Pairs of participants would discuss on the challenges and opportunities related to a particular theme during around twelve minutes using a push-through procedure. They added the output of their discussion using post-its to the discussion notes on the table. Each twelve minutes, the pairs progressed to the next theme at the same table and would add their findings to the already existing discussion notes. This round ended as soon as each pair of participants had addressed each theme once.

In the *synthesis round* the input from the different tables was grouped thematically, i.e. according to each of the five themes. First, the participants were asked to group the input (post-its) by grouping similar ideas. Next, participants were invited to individual dot-voting: each participant received eight stickers that he/she could use to highlight the most urgent or important challenges (4 stickers) and opportunities (four stickers). Additionally, each participant received two veto stickers to indicate their disagreement: one for an opportunity and one for a challenge. Participants were requested to put their initials on the veto sticker such that they could be prompted for more explanation during the plenary discussion round.

The goal of the *plenary discussion* was to use the input from the idea generation and synthesis rounds to define the most important challenges and opportunities of XLA for each of the five

themes. A list of most important challenges and opportunities was assembled based on the number of dot votes. Selected items for each of the themes were discussed one by one, and participants were invited to bring forward the findings and elaborate on them. Other participants were invited to comment and discuss. This discussion was recorded using audio recording devices.

Ethical approval was obtained from the ethical commission of KU Leuven. Workshop participants that subscribed to the workshop received an explanation of the research performed beforehand by email, including a notice that they would be invited to sign a consent form if they would participate in the plenary discussion. During the workshop, all participants were informed about the protocol and invited to sign the consent forms if they agreed their audio was recording during the plenary discussion. Participants not consenting could still participate in all parts of the workshop, except the final audio-recorded plenary discussion. At several stages during the workshop pictures were taken to support the processing of results. Prior consents was requested for participants being pictured.

3 RESULTS

The workshop organized at the EC-TEL 2019 conference in Delft, the Netherlands was attended by 44 participants. All attendants participated in the idea generation round. After a short break, 17 participants consented with the recording of their voices during the final discussion, and therefore participated in both the idea generation round, the synthesis round, and the plenary discussion. Below, we elaborate on the main opportunities and challenges that were identified during the synthesis round and the plenary discussion (as illustrated in Table 2), grouped by each of the five themes.

Table II: Most voted challenges and opportunities.

Theme	Item	#opportunity	#challenge	#veto
Opportunities				
Data	Course/learning design, re-using teachers' previous data	6	0	0
Stakeholders	Teachers reflect/understand own effectiveness (by visualization features)	4	1	0
Communication	Data storytelling	7	0	0
Evaluation	Evaluate impact of the explanations	3	1	0
Implementation & adoption	Adapt to the target groups	5	2	0
Challenges				
Data	Include how recent the data is	0	4	0
Stakeholders	Community building	1	4	0
Communication	Time-based LA, splitting explanation per phase/step	0	6	0
Evaluation	Added value for user (pre/post)?	1	3	0
Implementation & adoption	Support (technical, pedagogical) when system is deployed	1	3	0

3.1 Opportunity

Data. The main identified opportunity for data for XLA is to use data regarding course/learning design and reusing teachers' previous data (6 opportunity stickers, no challenge stickers, no veto).

During the discussion, participants elaborated that if humans generate the data, this might alleviate the explainability issue. After all, human-generated data could be useful to explain computer-generated data. The participants are interested to know how learning design, if well-modelled, will influence predictions and explanations and how feedback regarding these predictions and explanations can improve the underlying models. An example of such an improvement is elaborated on in the paper of Mothilal et al. (2018), where the explanatory technique of LIME is used to obtain explanations of the prediction of first-year engineering student success, which improved the model of student success. It is however still an open research challenge what data can and should be collected to support explainability. Nonetheless, the participants agreed that the expectations should be clear beforehand and that annotations on the data, if used with care, can support the explainability. A final warning regarding the data of course/learning design is that it can and should be context-specific, complicating wider use.

Stakeholders. The main identified opportunity of XLA for stakeholders within the workshop is for teachers when they can use XLA to reflect upon or understand their own effectiveness. (4 opportunity stickers, 1 challenge sticker, no veto). XLA will definitely provide an opportunity for teachers as XLA can disclose understandable explanations and recommendations to teachers regarding their teaching effectiveness. Participants also remarked that it would be a challenge however combine different perspectives of stakeholders, especially if they are conflicting. How should the perspectives be prioritized or weighted?

Communication. The main opportunity for the communication of XLA is data storytelling (7 opportunity stickers, 0 challenge stickers, no veto). Recent research and technological advancement have identified opportunities to automate data storytelling (Echeverria et al. 2018). It remains to be researched, however, to what level this automation is feasible and how and when a data scientist should still be in the loop. Storytelling, both manual and automatic, unlocks the opportunity of personalization. This immediately raises additional concerns related to ethics. For example, can personalized explanations trigger different interpretations depending on the personalization? Storytelling also has the opportunity to emphasize particular parts of the data, hereby providing an answer to the data abundance problem. One should be careful, however, not to 'obscure' the data: one should be transparent on which data is emphasized and which is hidden.

Evaluation. The main identified opportunity of the evaluation of XLA is related to the evaluation of the impact of the explanations (3 opportunity stickers, 1 challenge sticker, 0 veto). While the problem of evaluating XLA is new, there are consolidated techniques that could be adapted. Before evaluation can be started however, it is important to define clearly the different evaluation goals: they can range from perceived utility to impact on, e.g., advising and decision making. One should take care to not only set up separate evaluations with the different stakeholder groups, but to use the opportunity of mixed-group evaluations. The evaluation should moreover focus on both subjective and more objective indicators: one should not only rely on subjective statements but also attempt to look for objective/quantitative measures, such as the impact on learning gain.

Implementation & adoption. The main opportunity for the implementation and adoption of XLA is the adaptation to different target groups (5 opportunity stickers, 2 challenge stickers, no veto). Different target groups might need different types of explanations and interpretations of LA

predictions and recommendations. Each stakeholder might have particular needs and therefore, the explanations and interpretations should be personalized to the particular group of stakeholders.

3.2 Challenges

Data. The main identified challenge for data in XLA is to include information about and take into account how recent the data is (0 opportunity stickers, 4 challenge stickers, 0 veto stickers). It is challenging to find a threshold that could uniquely define what ‘recent’ and ‘old’ data are. There is a tension between how valuable old data (to obtain enough data to train the models or to show historical evolution) and new data (more representative of current state) is. Additionally, attention should be paid to how to explain to users which data is used in these models and that predictions and recommendations rely on past data. Finally, deploying XLA can, and most likely will, influence the data itself as it is expected to have an impact on actual learning and teaching.

Stakeholders. The main challenge for the stakeholders is to build a strong community. (1 opportunity sticker, 4 challenge stickers, no veto). For the entire field of LA it is a challenge to build a strong community that could support stakeholders working on explanations of predictions and recommendations en strengthen their collaboration and the adoption of XLA When done well, the explanations have the opportunity to foster trust among different stakeholders, for example among students and teachers in a MOOC.

Communication. The main challenge for communication within XLA is to consider the time dimension of learning analytics (0 opportunity stickers, 6 challenge stickers, 0 veto). Longitudinal data is challenging to handle within learning analytics. XLA should be able to provide explanations for the different phases over time. Moreover, these explanations should be tailored to the particular phases and contexts they are provided in.

Evaluation. The main challenge for a good evaluation of XLA is to identify the added value for the stakeholders (1 opportunity sticker, 3 challenge stickers, 0 veto). The evaluation of XLA should focus on identifying the added value of explanations for different stakeholders, and in particular should be able to show how the explanations contribute to what the stakeholders already know (e.g., using a pre/post test setup).

Implementation & adoption. The main challenge for the implementation and adoption of XLA are both technical and pedagogical support during deployment (1 opportunity sticker, 3 challenge stickers, 0 veto). The actual implementation and adoption of XLA will provide ample challenges, especially when deployments at scale are considered. These issues are not only of technical nature, but also pedagogical: how can the explanations be used appropriately during the learning process?

4 DISCUSSION AND CONCLUSION

This short paper called upon the input of more than 40 stakeholders to shape the domain of Explainable Learning Analytics (XLA), which aims at developing LA-specific advancement regarding Explainable Artificial Intelligence (XAI). In particular, this paper reports on the opportunities and challenges of XLA as identified by this group of international stakeholders collected during a 3-hour workshop at the 2019 European Conference on Technology Enhanced Learning (EC-TEL). Opportunities and challenges were collected regarding five main themes: data, stakeholders,

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

communication, evaluation, and implementation & adoption. The next step in future research would be to make a deeper analysis of the workshop outcomes and especially the audio recordings made and then to compare the outcomes to existent findings in XLA and XAI, for instance to the findings of Miller (2017) and Karga & Satratzemi (2019).

The input from stakeholders is undoubtedly valuable for the advancement of XLA. This contribution is, due to several limitations, only a small step towards a more profound integration of the different stakeholders in the development of the domain. A first limitation is that the workshop was held at the EC-TEL 2019 conference hereby causing a biased sample of the stakeholder population. The involved stakeholders were mainly researchers active in Technology Enhanced Learning, possibly causing the observed bias towards teachers in for instance the stakeholders dimension. Future stakeholder consultations should more heavily involve practitioners, policy-makers, and end-users of LA. Earlier work will provide inspiration regarding for instance the inclusion of students as a stakeholder in XLA (Putham & Conati 2019; Baria-Pineda & Brusilovsky 2019). A positive element of the stakeholder population is that they represented the wider domain of Technology Enhanced Learning, of which LA is only a sub-domain. On the negative side however, hereby introducing the second limitation, this meant that some attendants were not very acquainted with the specifics of the LA domain, while others were considered experts. The same holds for XAI: some attendants were experienced researchers or users, while others were not familiar with the domain. For future stakeholder consultations we recommend to set up a protocol that aims at better handling such differences in expertise, both regarding the LA and the XAI domain. A third limitation is the short duration of the workshop, which limited both the width and the depth of the discussion, and the limitations of the recordings made (only during the synthesis round).

To conclude, we can state that this paper contributes to the development of the XLA domain by the identification of challenges and opportunities regarding data, stakeholders, evaluation, communication, and implementation & adoption.

ACKNOWLEDGMENTS

We thank the participants of the EC-TEL 2019 XLA workshop for their contribution. This work was funded by the EU LALA project (grant no. 586120-EPP-1-2017-1-ES-EPPKA2-CBHE-JP). This project has been funded with support from the European Commission. This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

REFERENCES

- Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y., & Kankanhalli, M. (2018). Trends and trajectories for explainable, accountable and intelligible systems: An hci research agenda. In *Proceedings of the 2018 chi conference on human factors in computing systems* (pp. 582:1–582:18). New York, NY, USA: ACM. Retrieved from <http://doi.acm.org/10.1145/3173574.3174156> doi: 10.1145/3173574.3174156
- Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access*, 6, 52138–52160.

- Anjomshoae, S., Najjar, A., Calvaresi, D., & Frmling, K. (2019). Explainable agents and robots: Results from a systematic literature review. In (pp. 1078–1088).
- Barria-Pineda, J. & Brusilovsky, P. (2019) Making Educational Recommendations Transparent through a Fine-Grained Open Learner Model. IUI Workshops 2019
- Broos, T., Pinxten, M., Delporte, M., Verbert, K., & De Laet, T. (2019). Learning dashboards at scale: early warning and overall first year experience. *Assessment & Evaluation in Higher Education*, 1–20. <https://doi.org/10.1080/02602938.2019.1689546>
- Choo, J., & Liu, S. (2018). Visual analytics for explainable deep learning.
- Duval Erik (2012). Learning Analytics and Educational Data Mining, Erik Duval's Weblog, 30 January 2012, <https://erikduval.wordpress.com/2012/01/30/learning-analytics-and-educational-data-mining/>
- Echeverria, Vanessa and Martinez-Maldonado, Roberto and Granda, Roger and Chiluiza, Katherine and Conati, Cristina and Shum, Simon Buckingham (2018). Driving Data Storytelling from Learning Design. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK 2018) Sydney, Australia*, isbn: 978-1-4503-6400-3, (pp 131–140), doi: 10.1145/3170358.3170380
- Greller, Wolfgang & Drachsler, Hendrik. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics. *Educational Technology & Society*. 15. 42-57.
- Karga, S., & Satratzemi, M. (2019). Using explanations for recommender systems in learning design settings to enhance teachers' acceptance and perceived experience. *Education and Information Technologies*, 24(5), 2953–2974. <https://doi.org/10.1007/s10639-019-09909-z>
- Kwon, B. C., Choi, M.-J., Kim, J. T., Choi, E., Kim, Y. B., Kwon, S., ... Choo, J. (2019). Retainvis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records. *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 299–309.
- Miller, T. (2017). Explanation in Artificial Intelligence: Insights from the Social Sciences. *Artificial Intelligence*, 267, 1–38. <https://doi.org/arXiv:1706.07269v1>
- Mohseni, S., Zarei, N., & Ragan, E. (2018). A survey of evaluation methods and measures for interpretable machine learning. *arXiv.org*. Retrieved from <http://search.proquest.com/docview/2139809046/>
- Mothilal, R. K., & Broos, T. (2018). Predicting first-year engineering student success: from traditional statistics to machine learning. In *Proceedings of the 46th sefi conference* (pp. 1–8). Copenhagen, Denmark. Retrieved from <https://www.sefi.be/proceedings/?conference=copenhagen2018>
- Ouyang, S., Lawlor, A., Costa, F., & Dolog, P. (2018). Improving explainable recommendations with synthetic reviews.
- Putnam, V., & Conati, C. (2019). Exploring the Need for Explainable Artificial Intelligence (XAI) in Intelligent Tutoring Systems (ITS). IUI Workshops 2019
- Spinner, T., Schlegel, U., Schäfer, H., & El-Assady, M. (2019). explainer: A visual analytics framework for interactive and explainable machine learning. *IEEE transactions on visualization and computer graphics*.
- Vellido, A. (2019). The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural Computing and Applications*.
- Verbert, K., Duval, e., Klerkx, J., Govaerts, S., & Santos, J. (2013, 09). Learning analytics dashboard applications. *American Behavioral Scientist*, 57. doi: 10.1177/0002764213479363

- Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., & Klerkx, J. (2014). Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing*, 18(6), 1499–1514.
- Zhang, Q.-s., & Zhu, S.-c. (2018). Visual interpretability for deep learning: a survey. *Frontiers of Information Technology & Electronic Engineering*, 19(1), 27–39.
- Zhang, Y., & Chen, X. (2018). Explainable recommendation: A survey and new perspectives.
- Zhao, X., Wu, Y., Lee, D. L., & Cui, W. (2019). iforest: Interpreting random forests via visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 407–416.

Teacher-Centered Dashboards Design Process

Mohamed Ez-zaouia

University of Lyon, Université Jean Moulin Lyon 3, iaelyon school of Management
Woonoz, Lyon, France
ezzaouia.mohamed@gmail.com

ABSTRACT: Interest in dashboards in schools has been growing in recent years as they have great potential in fostering data transparency and informing teachers' practices. However, research surrounding them is not unified and even less transparent, because of a lack of guidance in grounding their design as a process tailored to end-users needs. We present a process model for teacher-centered dashboards as a design and validation process with four mutually informed stages: (1) situate the domain space by framing teachers' routines, practices, and needs, (2) ideate the domain into multiple alternatives and prototypes, (3) develop, and (4) evaluate the dashboard. We drive recommendations within each stage to inform the process. We borrowed the foundations of the model from research in the HCI and InfoVis fields. We apply our model to five case studies from literature. We find that this model can provide a framework to scaffold dashboards' design process, mutually inform the underlying stages, and guide consolidating artifacts. We reflect on our work to provide design implications to point towards explainable dashboards design to best support teachers.

Keywords: Teacher-centered design, Dashboards, Explainable dashboards, Design process

1. INTRODUCTION

Mainstream technology is becoming ubiquitous in today's classrooms (Technavio, 2016; Buabeng-Andoh, 2012; Millar, 2013). and, it has the potential to provide insightful information about the state of learning. Recently, out of 762 polled teachers, 86% think that data is important for being an effective teacher, and 81% think that students will benefit when data informs teaching (Campaign, 2018). Likewise, dashboards are becoming important assets to facilitate data transparency, sense-making, reflection, as well as sophisticated portals for teachers to inform their work, decisions, and practices (Verbert et al., 2013; Sedrakyan et al., 2019). Prior work has shown the potential of dashboards to assist teachers, for instance, in monitoring students' performance to mediate the classroom (Molenaar and van Campen, 2017), to allocate time to students of lower abilities (Holstein, McLaren, and Aleven, 2018), to plan lessons and debriefs (Xhakaj et al., 2017), and to provide personalized in-time support to students (Aslan et al., 2019).

Although the great potential and interdisciplinary opportunities for research, current literature on dashboards designed for teachers is, less unified, and less tailored towards end-users needs, which hampers the trust and adoption of such tools. For instance, a development that is gaining momentum in human-centered design, e.g., "participatory design", "design thinking", is still "under-presented" in technology-enhanced learning (McKenney and Kali, 2017). Moreover, information visualization (Infovis) techniques are not embraced well yet by visual learning analytics (Vieira et al., 2018). Additionally, there is still a lack of specific visualizations and visual metaphors addressing teachers' and students' unique needs in learning analytics dashboards (Schwendimann et al., 2017).

Furthermore, as pointed out by two recent systematic reviews (Schwendimann et al., 2017; Bodily et al., 2018), there is still a weakness in the process of design and evaluation of dashboards in learning contexts. We argue that this is primarily due to a lack of guidance in grounding the design as a transparent design process with a deep consideration of both technical and social aspects surrounding the design and evaluation of dashboards. Also, a lack of articulation in reporting not only on the end-artifacts but on the underlying design rationales and validations. This inspired the main motivation of our work: *Designers for teachers need models as systematic guiding principles to help scaffold the process of design and validation of dashboards to help foster trust and adoption.*

To inform the design of such interfaces, to foster transparency – through input from teachers, we present a process model for teacher-centered dashboards design with four mutually informed stages: (1) situate the domain space by framing teachers’ routines, practices, and needs, (2) ideate the framed domain into design goals, tasks, data, visual abstractions, design alternatives, and prototypes, (3) develop a dashboard, and (4) evaluate its significance by assessing teachers’ data-informed practices. We provide recommendations within each stage to inform design activities: exploring ideas, refining solutions, and consolidating artifacts.

By reflecting on our personal experience on the design and evaluation of dashboards for teachers, we turned to pertinent research in HCI and InfoVis to borrow the foundations of the model. HCI provides a wide range of methods that help empathize with teachers to understand their routines and needs (Wright and McCarthy, 2008), and to onboard them in a space of shared trust and knowledge while designing and validating interfaces (Muller and Kuhn, 1993). InfoVis, on the other hand, provides a tool-set to abstract a domain space into design goals (Lam et al., 2018), tasks (Brehmer and Munzner, 2013), and data semantics (Munzner, 2014) which are mapped into visual forms (Cleveland and McGill, 1984) to shape the interface of a dashboard. Together, HCI and InfoVis ensure a good fit between dashboards’ designs and the ways teachers are aiming to perform their everyday activities. We argue that designers not only need to be familiar with such approaches from both fields but they also need to appropriating (Dourish, 2003; Louridas, 1999) such approaches to scaffold a design process. To demonstrate this model, we apply it to concrete examples from literature through five case studies. We find that this model can provide a framework to structure dashboards’ design process, mutually inform underlying stages, help consolidate, and report on artifacts along the way.

2. BACKGROUND AND RELATED WORK

2.1. Informing Dashboard Design From HCI

The field of HCI provides a rich and varied range of methods to guide the design of dashboards. Roughly designerly practices fall into three main approaches: implicit, explicit, and process. Early work in design was more informed through implicit or primitive approaches. Bricolage is one primitive approach where designers try to blend different elements available in their immediate environment in making a new design (Louridas, 1999). Similar to bricolage, appropriation is another primitive approach that can be associated with customization in the sense that designers try to adapt, adopt, and reuse different elements into a new working design (Dourish, 2003). In both bricolage and appropriation, designers do not create new elements, but instead make use of existing ones (e.g.,

ideas, artifacts) to serve new purposes different than the ones for which the original design was intended.

Follow-up work informed the raise of explicit methods to provide formal guidelines as designers start developing professional ways of working with related formal education and qualification. One of the most current dominant explicit approaches is human-centered design (Bannon, 2011). That is, designers, before engaging in any design activity at their own, for instance, using bricolage or appropriation (Louridas, 1999; Dourish, 2003), first conduct a research design by engaging with end-users of the design, to develop a deep knowledge of their issues, needs, tasks, activities, and abilities. Based on the knowledge gained they design a solution, which they then evaluate with end-users, and iterate on the design as needed. It would be unsound and misleading to propose a valid dashboard solution based on inferring “end-users”. Although this approach helps overcome poor design, it might be challenging for designers to tailor the design for individual (or group of) people without making it less appropriate or even overwhelming for others. Other methods emerged with different add-ons to address this shortcoming. For instance, activity-centered design (Norman, 2006) aims at addressing this by not focusing on the interface as simply a means to perform some tasks but instead on the activities that the interface enables end-users in their everyday routines and practices. For instance, a dashboard for teachers may integrate several tasks such as tracking students’ idle moments, responses, rapid attempts, etc. but the main activity of teachers might be to mediate (or orchestrate) the classroom. Value sensitive design (Friedman, 1996) pushes this approach even further by focusing on human core values in designing the interface rather than tasks or activities. Returning to our example, in the context of value-sensitive design, we might devise the design of a dashboard as an “equalizer force” in a way to help a teacher ensuring equal progress to all students of the classroom.

Recently, design processes emerged to provide systematic heuristics to guide the activity of design. Design thinking (Brown, 2008) is gaining momentum in HCI research and industry, which is a set of hands-on methods to guide - iteratively, framing a problem (to solve) from wildly and diverse perspectives, critic and refine ideas to uncover an innovative solution that meets users’ needs. To this end, design thinking wraps three fundamental skills, namely, empathy, rapid prototyping, and empirical justifications. The first step is empathy (Wright and McCarthy, 2008; Beyer and Holtzblatt, 1999), advocating designers to immerse themselves in end-users’ lives to experience, first hand, their problems, contexts, and needs. Once designers frame a deep understanding of needs from end-users’ perspectives, the second step is to prototype by rapidly generating multiple approximations of design ideas to try and test with actual users as quickly as possible (Dow et al., 2011; B. Hartmann et al., 2006). The third step is to evaluate prototypes using empirical evidence to justify choices. Prototypes are not end-artifacts in themselves. Instead, they are used as concrete communicative proxies to seek both positive and negative feedback about how they impact certain users’ behaviors, reflect on design ideas, and learn insights to inform subsequent iterations. Another popular process is participatory design, rather than designing for people, advocates fundamentally designing with people by situating with them, to articulate their problems, identify their needs, and co-design solutions in cooperation (Muller and Kuhn, 1993). By doing so, the new design will directly support users’ skills, activities, and fit within their workplace. Participatory design and design thinking both build upon rapid prototyping and active collaboration with end-users. Each method either implicit, explicit or process has its strengths and weakness, and each will lead to different solutions and designs. As designers, we need

to appropriate them (Dourish, 2003; Louridas, 1999) to scaffold a design process to best support teachers' needs.

2.2. Informing Dashboard Design From InfoVis

InfoVis provides a wide range of methods to map domain problems and questions into visual forms and dashboards by capturing four fundamental elements: rationales, tasks, data, and visual encoding. Significant research has been devoted to guiding capturing, bridging between these abstractions, and explicitly describing them in formal ways (Munzner, 2014; Amar et al., 2005; Lam et al., 2018; Brehmer and Munzner, 2013; Carroll and Rosson, 2003). Articulating on the aforementioned HCI methods of knowing end-users' contexts, activities, and identifying their needs (Wright and McCarthy, 2008), designers need to produce an explicit representation of design goals (or rationales) in terms of claims about the aspects that the new design must address, and how every aspect impacts (enable/limit) specific end-users' behaviors (Carroll and Rosson, 2003). Next, designers need to translate domain-specific questions into task abstractions, such as identify extremes, analyze outliers, compare or retrieve values, etc. (Amar et al., 2005; Brehmer and Munzner, 2013). However, bridging between high-level questions and low-level tasks is a challenging endeavor. Goals analysis aims at addressing this by decomposing domain questions into immediate design goals (explore, describe, explain, confirm) before mapping them to concrete tasks (Lam et al., 2018). Formal tasks can be used then to facilitate data abstraction by describing properties of (related) data, namely real semantics (temporal, spatial, continuous, discrete, keys, values, dates), types (quantitative, ordinal, categorical), and datasets (table, graph, text, field, stream, static), as first-class objects that can be visualized (Munzner, 2014) by mapping such properties into visual forms (or visual encoding). Data abstraction is actually the method of effectively mapping data properties to both graphical elements and properties (Cleveland and McGill, 1984). Point, line, surface, and volume are the basic graphical elements that can be used and combined to create visual forms. Position, size, color, orientation, texture, and shape are graphical properties that can be used to decorate visual forms.

The essence of dashboards is to emphasize insightful indicators by compacting the needed (all related and relevant) information in a small amount of visual space to inform the audience in a meaningful, efficient, and actionable way (Few, 2006). Dashboards capitalize on human perceptual and cognitive abilities of processing visual information. As a result, they lay on visual design techniques for monitoring, exploration, presentation, communication, and storytelling to better address the needs of a target audience (Segel and Heer, 2010; Echeverria et al., 2018; Skau et al., 2015; Parsons, 2018; Kosara, 2016).

Informed by business data analytics, prior literature provides three roles for dashboards mainly, strategic, analytic and operational (Few, 2006; Smith, 2013; Sarikaya et al., 2018). We instead think that it is more beneficial and practical for dashboard design to directly build on techniques that have already developed and validated in InfoVis regarding interfaces design specifically, role of visualizations (e.g., exploratory, confirmatory, presentation) (Schulz et al., 2013), design goals (e.g., explore, describe, explain, confirm) (Lam et al., 2018), and analytical tasks (e.g., retrieve values, compare items, find extremum, filter, sort) (Amar et al., 2005; Brehmer and Munzner, 2013). Explicitly describing rationales, tasks, data, and visual encoding in "abstract" rather than domain-specific form, translates into three key benefits. First, it avoids oversimplifications and converging into local-

optimum solutions without exploring the design space of possibilities and alternatives. Second, it structures the validation of the newly designed artifact. And finally, it fosters transparency, trust, and adoption of the new design.

2.3. Engaging Teachers in the Design Loop

There is little research, mostly related to curricula (TEL) design, that has examined teachers as designers through different design processes. For instance, Roschelle and Penuel (2006), using a co-design approach, reported on dynamics and tensions between researchers, teachers, and developers in the following three phases (collecting requirements, rapid prototyping, software solidification) design process for (TEL) curriculum. Along the same line, Cober et al. (2015) highlight the vital role of teachers in participatory design. Some, on the other hand, are skeptical Kirschner (2015) about the approach of the teacher as a designer because they believe teachers - as professional - can adapt/adopt any TEL. They are not convinced by the benefit of engaging teachers in the design over the cost (time, resources, and energy) put in. We instead subscribe to the first call. That is, effective teachers are experts in the classroom's everyday routines (Hattie, 2012) thus having an essential role in bridging research, design, and practice.

2.4. Teachers' Dashboards Design Research

Unfortunately, the literature on design-based research and practices of design, analysis, and evaluation of teachers' dashboards is very scarce, with only very few exceptions. In two recent systematic reviews of more than 150 learning dashboards, almost half of the surveyed papers do not conduct any evaluation nor report on conducting a specific or using an existing design process (Schwendimann et al., 2017; Bodily et al., 2018). The first welcome exception is the framework proposed by Verbert et al. (2013) to guide the analysis of learning analytics dashboards. Although the framework is an excellent thinking tool, to evaluate the impact of a dashboard (e.g., see (Molenaar and van Campen, 2017), it mainly captures the evaluation part, and it does not provide a full model of how to design dashboards guiding the whole process from domain characterization to evaluation. Another welcome exception is the four stages workflow (problem identification, low-fidelity prototyping, high-fidelity prototyping, pilot studies) by Martinez-Maldonado et al. (2015) to guide the design and deployment of awareness tools for instructors and students. However, the workflow does not capture the principles of visualizations nor the challenges to tackle while designing dashboards. Our model aims at extending this latter by providing a process model built upon pertinent research from HCI and InfoVis.

3. APPROACH

This work is formed by reflecting on our experience in designing dashboards for teachers. In a process of introspection and analysis we generated, questioned, and interpreted practices surrounding dashboards' design, extensively reviewed literature from LAK, TEL, HCI, and InfoVis fields; we projected that understanding to articulate the conceptual model and, refined the reporting omitting evidence specific to our context. The process model that we describe in this paper is informed by previous models and methods aimed at applying visualization research to domain-specific problems. Namely, the model proposed by Munzner (2009) to guide and unify the analysis and validation of visualization tools through four nested levels, each with different threats of validity; for instance, in

the characterization level –of the domain space, the threat is “wrong problem” and validation is “observe and interview target users”. Although we find the nested model an excellent analysis tool, it does not provide a process approach of how to design nor offers practical advice to scaffold a design process. In fact, other models that build upon the nested model, have been proposed with the aim to provide a more holistic process. For example, a design study approach to conducting visualization research projects to solve a real-world problem through iterative stages (Sedlmair et al., 2012; McKenney and Kali, 2017). However, one main strand of these models is a lack of actionability. There persists a need for practicable models that do not compromise clarity and depth in the portrayal of the theoretical applicability. Our work is instead a process model offering a practical approach to devise design and validation of teachers’ dashboards by providing specific design knowledge within each stage guiding designer to explore, assess, and refine design alternatives and consolidate artifacts along the way. We refer to validation as an ongoing practice of justification of steps of the design and evaluation as the deployment of a dashboard for teachers in real-world settings. Finally, we aimed at applying the model to concrete examples from literature. This has the advantage to capitalize on previous thinking and research about dashboards’ design from multiple researchers and across different domains in the field which provide an initial, yet reliable validity of the model.

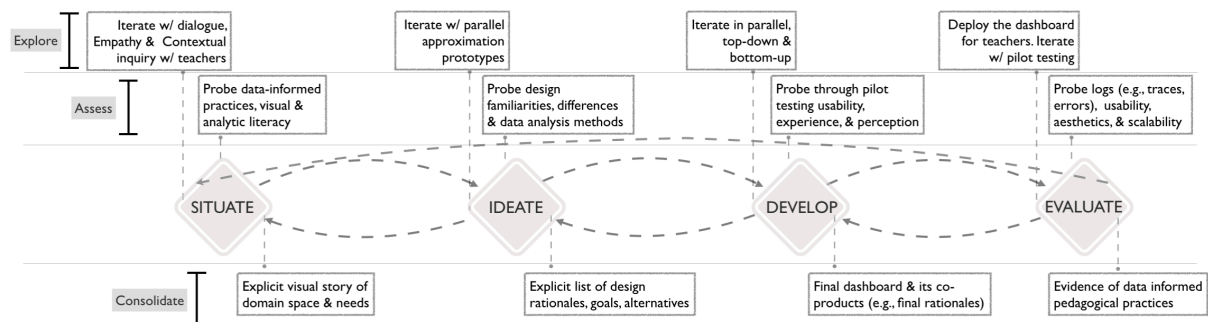


Figure 1: Four Stages of the model centered around three activities: explore, refine, consolidate.

4. PROCESS MODEL

Dashboard design is a process of solving a problem (Jonassen, 2000) to uncover a solution that meets users’ needs and giving it a form and shape. The process model assumes that we have already a problem, question, or idea to address through the design of a dashboard. Therefore, our model starts by situating the problem, exploring possible ideas of solutions, acting on those solutions by generating design approximations prototypes, assessing prototypes by seeking feedback and refining them before evaluating how they impact teachers’ practices and what behaviors they enable and limit in real-world settings (see Figure 1.).

4.1. Situate

Situate the domain space. Although the aim is to produce artifacts (e.g., dashboard), designers face phenomena whether facts (e.g., students’ progression), tasks (e.g., identify outliers), activities (e.g., class orchestration) or values (e.g., equal progress to students). These phenomena are situated and dynamic. They develop and change over time in specific places (e.g., classroom, school, home). When designing, we need to understand the interplay between a teacher and a dashboard through those phenomena, and other related entities (e.g., students, parents, staff members). Situating the domain

space is (1) capturing those phenomena and entities, (2) understanding their impact on teachers, and (3) explicitly describing their roles as parts and considerations of the design. Domain space reflects the possible range of motivations, needs, and constraints under which teachers are able and want to do their work in real-world settings.

Recommendations: The aim of this stage is to understand teachers' problems, tasks, activities, values, and needs in context. One way to achieve this is by conducting empirical research in real-world settings using to gather valuable insights. HCI provides a wide range of approaches to tackle this (see related work). Wright and McCarthy (2008) report the most commonly used methods in HCI to "know users" through dialogue and empathy (e.g., role play-based, scenario-based, interviews, observations). The contextual-inquiry methodology is another way to learn about how users perform tasks in context (Beyer and Holtzblatt, 1999). This method advocates immersing where users perform their activities to observe, ask questions, participate, empathize, and learn about users' practices, decisions, workflows, pain-points, constraints to gain insights and inspirations. At the end of this stage, as designers, we will be able to create an explicit description (visual story) of all the dimensions, phenomena, stakeholders that impact teachers in using the dashboard. This will help to form an explicit list of teachers' needs, i.e., all the claims that the design needs to accomplish.

4.2. Ideate

Ideate by generating multiple design ideas to address teachers' needs. Prior work, appropriation (Dourish, 2003), bricolage (Louridas, 1999) are common sources of inspiration and creativity. The goal is to explore multiple and wild design alternatives and to use various sources of evidence to learn insights, validate and refine solutions. Then act on those solutions by generating multiple design approximations prototypes (B. Hartmann et al., 2006). Parallel prototyping is an effective method to generate alternatives in parallel, which helps discover unseen constraints, local optimum, and new opportunities (Dow et al., 2011).

Recommendations: The goal of this stage is to iterate by creating multiple design prototypes to approximate solutions (Dow et al., 2011). Then to make prototypes tangible, e.g., paper frame, sketch, PowerPoint, wireframe so that we can externalize them early and often to seek feedback, and to validate design choices. While doing so, we need to create an explicit data abstraction to capture the types and attributes of data at hand, which will help consider all constraints early in the ideation phase. At some point in the ideation, prototyping needs to converge to "data sketching" by incorporating real data into digital prototypes. This will help discover unseen limitations, and gather practical insights. At the end of this stage, as designers, we will be able: to create an explicit description design rationale of the dashboard all the claims to accomplish with the design, to select main design alternatives prototypes (for development phase) with associated validations, and to create an explicit data abstraction and transformations (e.g., algorithm) to compute the needed indicators of the dashboard.

4.3. Develop

Within this stage, designers **develop** the validated design alternative prototypes to build a dashboard system. By the end of this stage, the dashboard needs to work using real data. To this end, designers need to address different design challenges to shape and put together all the required information on

the dashboard (see design challenges in related work). The key is to iterate in parallel using both top-down (user interface) and bottom-up (data, algorithm) to discover unseen constraints and/or and new opportunities early during the development. Javascript frameworks (e.g., React, Angular, Vue) and visualization libraries (e.g., D3js, Vega) can be useful to design interactive dashboards.

Recommendations: At this stage, it may be useful to iterate in parallel from both top-down by abstracting design goals, rationales, and needs into interactive visual encoding, views, pages, navigation, and layout to shape the dashboard, also from bottom-up by connecting dashboard with underlying data sources, and implementing the queries, algorithms, and transformations to expose indicators' data to the views of the dashboard. Adding traces loggers can be useful to gather insights into teachers' use of the dashboard (e.g., time spent, clicks). Adding audit loggers can also be useful to gather evidence on bugs and issues that may happen. Then, we need to conduct pilot testing to inspect if visual encodings meet teachers' needs, data, and visual literacy and whether there are other important details to consider. At the end of this stage, as designers, we will be able: to create an explicit description of the evolution of the interface of the dashboard through different iterations made – including iterations caused by constraints and/or and new emerged opportunities, to deploy the final concrete dashboard in real-world settings, and to create the final design rationales and needs as co-products of the dashboard.

4.4. Evaluate

At this stage, the goal is to **evaluate** the dashboard in teachers' context so that they can use it to inform their pedagogical practices. Evaluation often concerns larger-scale deployment and issues, and new needs will emerge. It may be useful to have a protocol of how to iterate on the dashboard during the period of the field deployment. Interestingly, at this stage, we loop back to the **situate** stage with the aim to understand the interplay between teachers and the newly designed dashboard considering all the phenomena (e.g., activities, tasks, motivations, pain-points) and entities (e.g., students, parents, home) to collect different source of evidence and learn insights about teachers' data-informed practices using the dashboard.

Recommendations: At this stage, we need to deploy the dashboard for teachers in real-world settings. We iterate first with pilot testings to inspect that the dashboard is working as expected and that the logs provide the needed insights into teachers' use of the dashboard. As this stage involves more teachers with diverse data, visual, and analytic literacy, it may be useful to inspect usability, aesthetics, and scalability issues as well as new needs that may emerge. At the end of this stage, we need be able to build an explicit description of teachers' overall use, experience, perception, pain-points, and suggestions as well as an explicit description of how the dashboard informed teachers' pedagogical practice (e.g. monitoring students, planning instructions, conducting lessons, providing debriefs, making sense of data, self-reflection).

5. ILLUSTRATIVE CASE STUDIES

5.1. Case Study 1

Ez-zaouia and Lavoué (2017) presented a teacher dashboard for the visualization of multi-modal students' emotions (Ez-zaouia and Lavoué, 2017). They propose an approach to use cloud APIs for

emotion recognition to infer students' emotions in online learning environments. They demonstrated their approach using a videoconferencing tool for foreign language training. Using audio and video streams they inferred automated emotions along with students' self-reported emotions. The dashboard presents a set of visualizations of students' discrete, dimensional and self-reported emotions.

The authors stated an interesting research question: "How can learners' inferred emotions be visualized by tutors to facilitate actionable feedback?". Using "how," we think that some qualitative work will be made to situate the problem and drive teachers' needs, before iterating, on ideas to uncover a solution. Instead, the authors select three data sources to infer emotions from, and they set four design principles for the dashboard. That is, the paper did not tackle the first two stages, situate and ideate, to motivate design rationales and goals concerning teachers' needs. However, the paper shines very well in the third stage (develop). The paper reports on the underlying architecture of the dashboard, an explicit description of data abstraction, and an extensive data analysis and transformations to drive the indicators of the dashboard. The paper also reports well on the visual abstraction (or encoding) for all the visualizations of the dashboard. For example, describing that they used a bubble and star as markers for discrete and dimensional emotion, and the size of both bubbles and stars is mapped to a derived score of emotions. However, task abstraction is not addressed. The authors conducted a pilot study using a questionnaire with two teachers. Finally, the dashboard was not deployed in real-world settings for evaluation.

5.2. Case Study 2

Ruiz et al. (2016) presented both students and teacher dashboard for the visualization of students' emotions (Ruiz et al., 2016). They propose an approach to use self-reported emotions to infer students' emotions in online learning environments. They first used google docs to validate both a questionnaire and prototypes of visualizations. Then they integrated the questionnaire and visualizations in a tool used by students. Students are then asked to fill the questionnaire before and after the class by reflecting on their emotions. Students' responses to the questionnaire are used to feed the visualizations, which students and teacher have access to.

The authors stated an interesting research question: "How can students' emotions be visualized to promote self-reflection?". Here also we find no qualitative work made to situate the problem and drive students and teachers needs in terms of what are the appropriate ways to enable students to express their emotions (e.g., questions, emojis, text, drawings, photos), and the appropriate ways to reflect back such information to students and teachers to enable self-reflection. In the develop stage, the paper reports very well on the visual abstraction, stating for instance that stacked bars are mapped to the average rating of every student's emotion for all sessions versus the average ratings of the group. However, data and task abstractions are not addressed in the paper. The authors conducted different iterations on the design of the dashboard before being integrated into a learning application called PresenceClick, but without end-user-driven justifications. The paper excels in the evaluation stage; the authors deployed the dashboard for students and teachers in real-world settings evaluating the usability, usefulness, and impact of the dashboard on mainly, students' motivation using logs and satisfaction questionnaires.

5.3. Case Study 3

Fu et al. (2017) presented both students and teacher a dashboard for the visualization of students' difficulties and differences while learning the C programming language in the classroom. They propose an approach to collect students' learning logs from a learning tool called BookLooper to feed the dashboard, which was integrated (as a plugin) into Moodle.

The paper reports on the develop stage describing in great detail the data abstraction, but not tasks nor the transformations to drive the indicators of nine sophisticated visualizations that shape the dashboard. The authors reported on the visual abstraction, for example, regarding a heatmap, they stated that the color of the cells encodes the number of times students try to compile C programs. However, justification of the choices made regarding all the visualizations are subjective to authors themselves (e.g., "With this heat-map chart, we can easily detect the activity and inactivity of students"). That is, the three other stages, situate to understand teachers' and students' problems and gather evidence about their needs, ideate to uncover the solution that meets the needs, then evaluate are not tackled by the paper.

5.4. Case Study 4

Gruzd and Conroy (2018) presented a dashboard for the visualization of students' interactions with learning materials resources and their fellow students in the class. They propose an approach to collect logs about students' discussions on Twitter, which is used by instructors for both formal and informal teaching.

The authors stated a research question to address by the design: "What analytical techniques would instructors like to see in a LA dashboard to support their assessment of Twitter facilitated discussions?". In fact, in the situate stage, they used both qualitative and qualitative evidence to understand instructors' needs using a survey administrated to 54 higher education instructors. Then, the authors analyzed instructors' responses to extract needs, which they then used to inform the design of the visualizations of a prototype dashboard. The authors reported on the visual encoding and few rationales behind their choice based on both related work and evidence from the questionnaire. However, there was no ideation to explore the space of possibilities and alternatives. Similarly, the authors did not report on the data and tasks abstractions nor tackled the develop and the evaluate stages.

5.5. Case Study 5

Holstein, Hong, et al. (2018) presented a virtual reality glasses dashboard for the visualization of real-time student performance indicators using an intelligent tutoring system in the classroom. In the situate stage, the authors based their work design findings from a previous work they have conducted to gather teachers' needs where teachers converged towards the idea of using eyeglasses giving them access to different indicators about students' performance indicators. During the ideation stage, they first conducted an in-lab storyboarding, brainstorming, and lo-fi prototyping with three teachers using papers-sketches, photoshop, and a combination of plastic eyeglasses and images on a computer to simulate the classroom. The first lo-fi prototyping session revealed that it was difficult for the teacher to embrace an actual class using mixed-reality glasses. In the next sessions, the authors used real

smart glasses. After, the authors moved to mid-fi prototyping. Next, the authors used contextual design and affinity diagramming by analyzing interviews and think-aloud sessions data to extract an explicit list of design rationales. The authors did not report on the data and task abstractions nor the visual encoding or design alternatives of the views of the dashboard. In the develop stage, the authors designed a hi-fi prototype that was used to conduct 10 sessions with teachers in simulated classrooms where the authors iterated on the design based on teachers' feedback. The tool was deployed during a single session for a pilot evaluation.

6. DISCUSSION AND DESIGN IMPLICATIONS

6.1. Reflecting on the Process Model

Our model does not attempt to be a full teacher inclusive design process. Several steps can be assessed thought pilot testing before validation with teachers. However, we believe that the situate stage is crucial to engage with teachers to understand their problems, contexts, needs, goals, values, and suggestions. Our work by no means attempts to propose a unique design process model for teachers' dashboards. Instead, we aim to articulate a process model to help inform and structure the design for teachers as a design process. Given the impact that a dashboard might have, not only on teachers, but also on students, parents, and other stakeholders, we argue that models that are more specific need to be proposed, implemented and tested.

Current literature of learning analytics dashboards (Bodily et al., 2018; Schwendimann et al., 2017) conveys the results of proceeding directly to the develop stage without much characterization of the interplay between a teacher and a dashboard in real-world settings. We echo that this field will benefit as much as from experiences in characterizing teachers' routines, practices, and particular types of problems they face and how data and dashboards can address them. Our case studies show that dashboards do not build upon explicit tasks nor design goal abstractions. Lack of such abstractions makes it difficult to conduct systematic performance evaluations among different dashboards (Schwendimann et al., 2017). This highlights a need for more focused design-based models and principles to guide dashboards' design, analysis, and validation.

Our process model sits between an analysis model and a systematic design process. We describe four stages of designing teachers' dashboards with steps and recommendations within each stage. We do not aim at providing fully structured (holistic) directives to design a dashboard. Instead, we aim at a flexible model of how to explore, refine, make and report on artifacts in design and use of teachers' dashboards, thus supporting designers to appropriate (Dourish, 2003) the four stages of our model as building design-blocks to devise and scaffold their process regarding their own needs, contexts, and constraints.

6.2. Designing for and with Teachers

Reflecting on our personal experience, we echo four implications of our model in designing for teachers. Although there is a similarity between these four challenges, they are neither completely independent nor equivalent. Their main distinction lays in the stage where they unfold, so we must consider them separately.

Design for Diverse and Situated Needs. Designers for professionals count on a consensus in users' needs when framing domain space, and abstracting it into an interface. However, teachers have a complex and changing context, different workflows and practices. They have different interests in using a dashboard to achieve different outcomes, which may be challenging to address through a fixed design (Sarikaya et al., 2018; Schwendimann et al., 2017).

Design for Different Data, Visual and Analytic Literacy. Designers for professionals build upon a homogeneity among users' visual literacy. However, such homogeneity is scarce among teachers, and they have different visual and analytic literacy, which need to be addressed using tailored representations (Sarikaya et al., 2018). We have been amazed to know that some teachers rely on their colleagues to manage tools to inform their practices. Others recommended sophisticated interactions such as sort, hide, resize from tools like a spreadsheet (Barbara Wasson, 2015).

Design for Robustness. Professionals can adjust to perform the task with the interface at hand. However, Teachers have very limited resilience to new interfaces, especially if they find it incomplete for their own needs, their way of doing things, and their familiarity with other interfaces.

Design for Attractiveness. Professionals are intrinsically motivated to use dashboards to perform their work. Teachers attempt to prefer instruction over formative assessment. Some might think that spending an hour on a dashboard to formally inform their practices is an hour wasted where they could be instructing students. Addressing both usability and aesthetic (J. Hartmann et al., 2007) will support teachers' adoption of the dashboard.

6.3. Towards Transparent Teachers' Dashboards

Although dashboards may have a beneficial story to positively empower teachers, they entail different challenges distinguishing between: social, cognitive, and technical. First, dashboards by their nature aim at capturing, summarizing, and presenting a set of measurable indicators. However, other important metrics are often omitted during both the design and evaluation of dashboards, which can be done on purpose, as such metrics are hard to quantify, e.g., teachers' experience, perception, pain-points, and frustrations using dashboards. We suggest that dashboards for teachers are deeply embodied in rich and diverse socio-cultural practices that although hard to observe, quantify and integrate, might provide valuable insights to inform the design and evaluation of dashboards, to best support teachers.

Similarly, teachers' reliance on and trust in dashboards are important factors to quantify. This is important as (black-box) AI or machine learning now powers several learning dashboards, where uncertain or even inaccurate inferences can be made, which may lead to inappropriate interpretations. Besides, teachers are often confronted with the black box and sophisticated nature of dashboards, and the associated learning platforms, which may hamper their trust in dashboards. Although, how to best design dashboards to assist teachers in developing informed strategies so that such systems empower their judgment in context and in a way to hinder over-reliance, and foster trust in the long term is still to be explored. Additionally, dashboards build upon the notion of data collection, processing, sampling, and selection of a subset of metrics to visualize to inform the audience. Even when this process is properly conducted to compute accurate information, metrics on dashboards can be misinterpreted by teachers (Barbara Wasson, 2015), for instance, depending on

their data, visual, and analytic literacy. Besides, the process itself might lead to losing the variation of data through summarization, or even reducing the quality of data, in both cases the interpretation of a dashboard may lead to inappropriate decisions and biases. Finally, dashboards rely on collecting, storing, and processing data. Surprisingly, ethics and privacy was not a major concern of many dashboards papers that we surveyed, except two papers that explicitly highlighted ethical concerns regarding learners' tracking (Ruiz et al., 2016) and transparency of the underlying technology of learning analytics (Aslan et al., 2019), and both papers were dealing with emotional information tracking. Ethics and privacy concerns should be addressed to provide enough information, to different stakeholders, regarding the collection, use, and design of data in dashboards.

6.4. Towards Explainable Roles of Teachers' Dashboards

We articulate five roles of dashboards with some underlying design considerations (DC). We aim therefore at abstracting dashboards' ill-defined (complex) goals (e.g., "monitoring", "exploration"), into low-level tasks (e.g., "validate indicators", "discover insights"), then into explicit considerations (DC), to guide designers in leveraging the desired information (e.g., indicators), while considering the task and purpose of each view on the dashboard.

Monitoring – Validate Indicators. Monitoring dashboards require close attention from the target audience to validate indicators related to data. Thus, their design needs to (DC1) allow a user to keep an eye on events that are in constant change, using (DC2) a reasonable data refresh rate, and (DC3) providing formative, quality, and safety ensuring metrics. Additionally, the design need to (DC4) grab users' attention immediately if any monitored indicators become invalidated, and (DC5) allow users to take immediate action.

Exploration – Discover Insights. Exploratory dashboards require direct manipulation and sense-making from the target audience to discover insights about data. Thus, their design needs to (DC1) provide different perspectives of data often using different views, (DC2) allows a user to manipulate and change different parameters related to data and (DC3) switch between different views. Finally, the design need to (DC4) allows a user to generate, ask, and interpret different questions about the data.

Presentation – Confirm Facts. Presentative dashboards require a glance view from the target audience to confirm (specific) facts about data. Their design needs to (DC1) be explanatory to educate and/or inform a user, (DC2) be augmented through annotations to create a long-lasting impression, and (DC3) enable memorability, engagement, and learnability. Further, the design is (DC4) often specific and compact rather than general and scalable (Kosara, 2016).

Communication – Convey Messages. Communicative dashboards require a glance view from a target audience, but in contrast to presentation techniques, the aim is to convey (multifaceted) messages rather than presenting a set of information. Thus, their design needs to (DC1) often address to a wide range audience with different (or even low) visual literacy and hence (DC2) build upon an ill characterization of the audience. Additionally, the design needs to (DC3) uses different embellishments in charts using domain-specific knowledge and metaphor to communicate the message while (DC4) avoiding distraction from the pure visual representations of data (Skau et al., 2015; Parsons, 2018).

Storytelling – Persuade Users. Storytelling dashboards require a glance view from the target audience, but in contrast to both presentation and communication techniques, the aim is to persuade users of some facts through data. Thus, their design needs to (DC1) help the user reason about those facts by providing arguments, (DC2) use specific interactions to sequence those arguments. Additionally, the design need to (DC3) combines data-driven indicators with textual contexts in a narrative way to create and tell a story (Segel and Heer, 2010; Echeverria et al., 2018).

These techniques, however, are neither completely independent nor equivalent nor mutually exclusive. Designers need to appropriate all of them to address different needs and visual literacy. Thus, understanding the strengths and weaknesses of each technique will help choose, combine and augment these techniques to craft a dashboard. For instance, connected scatter plot, cloud-words, sankey, stream graph, treemap, bubble-chart are some graphics that are known to work well for presentation (Kosara, 2016). Similarly, isotype, domain-specific graphics, glyphs, and more general metaphors are known to work well for communication (Skau et al., 2015). However, techniques such as presentation or communication might not be appropriate for exploration where the purpose is to support sense-making.

7. LIMITATIONS AND FUTURE WORK

The main limitation of the current work is the lack of “active” empirical evidence by applying the model to a concrete dashboard design to illustrate and support its validity. While applying the model to research from literature provides an initial validity of the model, we hope to take it to implementation and research to design, analyze, and validate dashboards in future work.

8. CONCLUSION

In this paper, we presented a process model for teacher-centered dashboards design. We articulated our model by reflecting on our personal experiences along with an expanded literature review from LAK, TEL, InfoVis, and HCI research. Our model articulates four mutually informed stages: situate, ideate, develop, and evaluate. We demonstrated our model through five case studies from the literature. We found that our model can provide a framework to structure dashboards’ design process, mutually inform underlying stages, guide consolidate, and report on artifacts along the way. We provide design implications to support teachers’ dashboards design. We hope our work provides a new perspective on teachers’ design, highlights its value and research.

REFERENCES

- Amar, R., Eagan, J., and Stasko, J. (2005). Low-level components of analytic activity in information visualization. IEEE Symposium on Information Visualization, 2005. INFOVIS 2005. <https://doi.org/10.1109/infvis.2005.1532136>
- Aslan, S., Alyuz, N., Tanriover, C., Mete, S. E., Okur, E., DMello, S. K., and Esme, A. A. (2019). Investigating the impact of a real-time, multimodal student engagement analytics technology in authentic classrooms. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI 19. <https://doi.org/10.1145/3290605.3300534>

- Bannon, L. (2011). Reimagining HCI: Toward a more human-centered perspective. *Interactions*, 18(4), 50–57. <https://doi.org/10.1145/1978822.1978833>
- Barbara Wasson, C. H. (2015). Data literacy and use for teaching (pp. 56–40). Taylor and Francis. <https://doi.org/10.4324/9781315777979>
- Beyer, H., and Holtzblatt, K. (1999). Contextual design. *Interactions*, 6(1), 32–42. <https://doi.org/10.1145/291224.291229>
- Black, P., and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7–74.
- Bodily, R., Kay, J., Aleven, V., Jivet, I., Davis, D., Xhakaj, F., and Verbert, K. (2018). Open learner models and learning analytics dashboards: A systematic review. *Proceedings of the 8th International Conference on Learning Analytics & Knowledge*, 41–50. <https://doi.org/10.1145/3170358.3170409>
- Brehmer, M., and Munzner, T. (2013). A multi-level typology of abstract visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2376–2385. <https://doi.org/10.1109/tvcg.2013.124>
- Brown, T. (2008). Design thinking. *Harvard business review*, 86(6), 84.
- Buabeng-Andoh, C. (2012). Factors influencing teachers' adoption and integration of information and communication technology into teaching: A review of the literature. *International Journal of Education and Development Using Information and Communication Technology (IJEDICT)*, 8, 136–155. <https://doi.org/10.1080/09523987.2018.1439712>
- Campaign, D. Q. (2018). What parents and teachers think about education data. <https://dataqualitycampaign.org/resource/what-parents-and-teachers-think-about-education-data/>
- Carroll, J. M., and Rosson, M. B. (2003). CHAPTER 15—Design rationale as theory. In J. M. Carroll (Ed.), *HCI models, theories, and frameworks* (pp. 431–461). Morgan Kaufmann. <https://doi.org/10.1016/B978-155860808-5/50015-0>
- Cleveland, W. S., and McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387), 531. <https://doi.org/10.2307/2288400>
- Cober, R., Tan, E., Slotta, J., So, H.-J., and Könings, K. D. (2015). Teachers as participatory designers: Two case studies with technology-enhanced learning environments. *Instructional Science*, 43(2), 203–228. <https://doi.org/10.1007/s11251-014-9339-0>
- Dourish, P. (2003). The appropriation of interactive technologies: Some lessons from placeless documents. *Computer Supported Cooperative Work (CSCW)*, 12(4), 465–490. <https://doi.org/10.1023/a:1026149119426>
- Dow, S. P., Glassco, A., Kass, J., Schwarz, M., Schwartz, D. L., and Klemmer, S. R. (2011). Parallel prototyping leads to better design results, more divergence, and increased self-efficacy. In *Design thinking research* (pp. 127–153). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-21643-5_8
- Echeverria, V., Martinez-Maldonado, R., Granda, R., Chiluiza, K., Conati, C., and Shum, S. B. (2018). Driving data storytelling from learning design. *Proceedings of the 8th International Conference on Learning Analytics & Knowledge*, 131–140. <https://doi.org/10.1145/3170358.3170380>

- Ez-zaouia, M., and Lavoué, E. (2017). EMODA: A tutor oriented multimodal and contextual emotional dashboard. *Proceedings of the Seventh International Learning Analytics and Knowledge Conference*, 429–438. <https://doi.org/10.1145/3027385.3027434>
- Few, S. (2006). Information dashboard design.
- Flórez-Aristizábal, L., Cano, S., Collazos, C. A., Solano, A. F., and Brewster, S. (2019). DesignABILITY: Framework for the design of accessible interactive tools to support teaching to children with disabilities. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI 19*. <https://doi.org/10.1145/3290605.3300240>
- Friedman, B. (1996). Value-sensitive design. *Interactions*, 3(6), 16–23. <https://doi.org/10.1145/242485.242493>
- Fu, X., Shimada, A., Ogata, H., Taniguchi, Y., and Suehiro, D. (2017). Real-time learning analytics for programming language courses. *Proceedings of the Seventh International Learning Analytics and Knowledge Conference*, 280–288. <https://doi.org/10.1145/3027385.3027407>
- Gruzd, A., and Conroy, N. (2018). Designing a learning analytics dashboard for twitter-facilitated teaching. *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, 46:1-46:4. <https://doi.org/10.1145/3231644.3231704>
- Gustafson, K. L., and Branch, R. M. (1997). Revisioning models of instructional development. *Educational Technology Research and Development*, 45(3), 73–89. <https://doi.org/10.1007/bf02299731>
- Hartmann, B., Klemmer, S. R., Bernstein, M., Abdulla, L., Burr, B., Robinson-Mosher, A., and Gee, J. (2006). Reflective physical prototyping through integrated design, test, and analysis. *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology - UIST 06*. <https://doi.org/10.1145/1166253.1166300>
- Hartmann, J., Sutcliffe, A., and De Angeli, A. (2007). Investigating attractiveness in web user interfaces. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 387–396. <https://doi.org/10.1145/1240624.1240687>
- Hattie, J. (2012). *Visible learning for teachers*. Routledge. <https://doi.org/10.4324/9780203181522>
- Holstein, K., Hong, G., Tegene, M., McLaren, B. M., and Aleven, V. (2018). The classroom as a dashboard: Co-designing wearable cognitive augmentation for k-12 teachers. *Proceedings of the 8th International Conference on Learning Analytics & Knowledge*, 79–88. <https://doi.org/10.1145/3170358.3170377>
- Holstein, K., McLaren, B. M., and Aleven, V. (2018). Student learning benefits of a mixed-reality teacher awareness tool in AI-Enhanced classrooms. In *Lecture notes in computer science* (pp. 154–168). Springer International Publishing. https://doi.org/10.1007/978-3-319-93843-1_12
- Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, 48(4), 63–85. <https://doi.org/10.1007/bf02300500>
- Kirschner, P. A. (2015). Do we need teachers as designers of technology enhanced learning? *Instructional Science*, 43(2), 309–322. <https://doi.org/10.1007/s11251-015-9346-9>
- Kosara, R. (2016). Presentation-oriented visualization techniques. *IEEE Computer Graphics and Applications*, 36(1), 80–85. <https://doi.org/10.1109/mcg.2016.2>
- Lam, H., Tory, M., and Munzner, T. (2018). Bridging from goals to tasks with design study analysis reports. *IEEE Transactions on Visualization and Computer Graphics*, 24(1), 435–445. <https://doi.org/10.1109/tvcg.2017.2744319>

- Louridas, P. (1999). Design as bricolage: Anthropology meets design thinking. *Design Studies*, 20(6), 517–535. [https://doi.org/10.1016/s0142-694x\(98\)00044-1](https://doi.org/10.1016/s0142-694x(98)00044-1)
- Maldonado, R. M., Kay, J., Yacef, K., and Schwendimann, B. (2012). An interactive teacher's dashboard for monitoring groups in a multi-tabletop learning environment. In *Intelligent tutoring systems* (pp. 482–492). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30950-2_62
- Martinez-Maldonado, R., Pardo, A., Mirriahi, N., Yacef, K., Kay, J., and Clayphan, A. (2015). LATUX: An iterative workflow for designing, validating, and deploying learning analytics visualizations. *Journal of Learning Analytics*, 2(3), 9–39. <https://doi.org/10.18608/jla.2015.23.3>
- McKenney, S., and Kali, Y. (2017). Design methods for TEL. In *Technology enhanced learning* (pp. 37–46). Springer International Publishing. https://doi.org/10.1007/978-3-319-02600-8_4
- Millar, E. (2013). How teachers and students are adjusting to the digital classroom. <https://www.theglobeandmail.com/life/teachers-and-students-adjust-to-the-digital-classroom/article14012464/>
- Molenaar, I., and Campen, C. K. (2017). Teacher dashboards in practice: Usage and impact. In *Data driven approaches in digital education* (pp. 125–138). Springer International Publishing. https://doi.org/10.1007/978-3-319-66610-5_10
- Muller, M. J., and Kuhn, S. (1993). Participatory design. *Commun. ACM*, 36(6), 24–28. <https://doi.org/10.1145/153571.255960>
- Munzner, T. (2009). A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics*, 15(6), 921–928. <https://doi.org/10.1109/tvcg.2009.111>
- Munzner, T. (2014). What: Data abstraction (pp. 20–40). A K Peters/CRC Press. <https://doi.org/10.1201/b17511>
- Norman, D. A. (1991). Cognitive artifacts. *Designing Interaction: Psychology at the Human-Computer Interface*, 1, 17–38.
- Norman, D. A. (2006). Logic versus usage: The case for activity-centered design. *Interactions*, 13(6), 45–ff. <https://doi.org/10.1145/1167948.1167978>
- Parsons, P. (2018, October). Conceptual metaphor theory as a foundation for communicative visualization design.
- Roschelle, J., and Penuel, W. R. (2006). Co-design of innovations with teachers: Definition and dynamics. *Proceedings of the 7th International Conference on Learning Sciences*, 606–612. <http://dl.acm.org/doi/10.1145/1150034.1150122>
- Ruiz, S., Charleer, S., Urretavizcaya, M., Klerkx, J., Fernández-Castro, I., and Duval, E. (2016). Supporting learning by considering emotions: Tracking and visualization a case study. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 254–263. <https://doi.org/10.1145/2883851.2883888>
- Sarikaya, A., Gleicher, M., and Szafir, D. A. (2018). Design factors for summary visualization in visual analytics. *Computer Graphics Forum*, 37(3), 145–156. <https://doi.org/10.1111/cgf.13408>
- Schulz, H.-J., Nocke, T., Heitzler, M., and Schumann, H. (2013). A design space of visualization tasks. *IEEE Transactions on Visualization and Computer Graphics*, 19(12), 2366–2375. <https://doi.org/10.1109/tvcg.2013.120>
- Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., and Dillenbourg, P. (2017). Perceiving learning at a glance: A systematic literature

- review of learning dashboard research. *IEEE Transactions on Learning Technologies*, 10(1), 30–41. <https://doi.org/10.1109/tlt.2016.2599522>
- Sedlmair, M., Meyer, M., and Munzner, T. (2012). Design study methodology: Reflections from the trenches and the stacks. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2431–2440. <https://doi.org/10.1109/tvcg.2012.213>
- Sedrakyan, G., Mannens, E., and Verbert, K. (2019). Guiding the choice of learning dashboard visualizations: Linking dashboard design and data visualization concepts. *Journal of Computer Languages*, 50, 19–38. <https://doi.org/10.1016/j.jvlc.2018.11.002>
- Segel, E., and Heer, J. (2010). Narrative visualization: Telling stories with data. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), 1139–1148. <https://doi.org/10.1109/tvcg.2010.179>
- Shneiderman, B. (2003). The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization* (pp. 364–371). Elsevier. <https://doi.org/10.1016/b978-155860915-0/50046-9>
- Smith, V. S. (2013). Data dashboard as evaluation and research communication tool. *New Directions for Evaluation*, 2013(140), 21–45. <https://doi.org/10.1002/ev.20072>
- Technavio. (2016). Digital classroom market will grow at an impressive CAGR of almost 13% until 2020. <https://www.businesswire.com/news/home/20161208005416/en/Digital-Classroom-Market-Grow-Impressive-CAGR-13>
- Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Assche, F. V., Parra, G., and Klerkx, J. (2013). Learning dashboards: An overview and future research opportunities. *Personal and Ubiquitous Computing*. <https://doi.org/10.1007/s00779-013-0751-2>
- Vieira, C., Parsons, P., and Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers and Education*, 122, 119–135. <https://doi.org/10.1016/j.compedu.2018.03.018>
- Wright, P., and McCarthy, J. (2008). Empathy and experience in HCI. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 637–646. <https://doi.org/10.1145/1357054.1357156>
- Xhakaj, F., Aleven, V., and McLaren, B. M. (2017). Effects of a dashboard for an intelligent tutoring system on teacher knowledge, lesson plans and class sessions. In *Lecture notes in computer science* (pp. 582–585). Springer International Publishing. https://doi.org/10.1007/978-3-319-61425-0_69

CrossMMLA in practice: Collecting, annotating and analyzing multimodal data across spaces

Michail Giannakos, NTNU, Norway; Daniel Spikol, University of Malmö, Sweden; Inge Molenaar, Radboud University, The Netherlands; Daniele Di Mitri, Open University, The Netherlands; Kshitij Sharma, NTNU, Norway; Xavier Ochoa, New York University, NY, USA; Rawad Hammad, University of East London, UK

ABSTRACT: Learning is a complex process that is associated with many aspects of interaction and cognition (e.g., hard mental operations, cognitive friction etc.) and that can take across diverse contexts (online, classrooms, labs, maker spaces, etc.). The complexity of this process and its environments means that it is likely that no single data modality can paint a complete picture of the learning experience, requiring multiple data streams from different sources and times to complement each other. The need to understand and improve learning that occurs in ever increasingly open, distributed, subject-specific and ubiquitous scenarios, require the development of multimodal and multisystem learning analytics. Following the tradition of CrossMMLA workshop series, the proposed workshop aims to serve as a place to learn about the latest advances in the design, implementation and adoption of systems that take into account the different modalities of human learning and the diverse settings in which it takes place. Apart from the necessary interchange of ideas, it is also the objective of this workshop to develop critical discussion, debate and co-development of ideas for advancing the state-of-the-art in CrossMMLA.

Keywords: multimodal learning analytics, learning spaces, sensor data

1 BACKGROUND

The field of multimodal learning analytics (MMLA) is an emerging domain of Learning Analytics and plays an important role in expanding Learning Analytics goal of understanding and improving learning in all the different environments where it occurs. The challenge for research and practice in this field is how to develop theories about the analysis of human behaviors during diverse learning processes and to create useful tools that could that augment the capabilities of learners and instructors in a way that is ethical and sustainable. CrossMMLA workshop will serve as a forum to exchange ideas on how we can analyze evidence from multimodal and multisystem data and how we can extract meaning from these increasingly fluid and complex data coming from different kinds of transformative learning situations and how to best feedback the results of these analyses to achieve positive transformative actions of those learning processes. CrossMMLA aims at helping learning analytics to capture students' learning experiences across diverse learning spaces. The challenge is to capture those interactions in a meaningful way that can be translated into actionable insights (e.g., real-time formative assessment, post reflective reviews; Di Mitri et al., 2018, Echeverria et al., 2019) .

MMLA uses the advances in machine learning and affordable sensor technologies (Ochoa, 2017) to act as a virtual observer/analyst of learning activities. Additionally, this virtual nature allows MMLA to provide new insights into learning processes that happen across multiple contexts between

stakeholders, devices and resources (both physical and digital), which often are hard to model and orchestrate (Scherer et al., 2012; Prieto et al., 2018). Using such technologies in combination with machine learning, LA researchers can now perform text, speech, handwriting, sketches, gesture, affective, or eye-gaze analysis (Donnelly et al., 2016; Blikstein & Worsley, 2016, Spikol et al., 2018), improve the accuracy of their predictions and learned models (Giannakos et al., 2019) and provide automated feedback to enable learner self-reflection (Ochoa et al, 2018). However, with this increased complexity in data, new challenges also arise. Conducting the data gathering, pre-processing, analysis, annotation and sense-making, in a way that is meaningful for learning scientists and other stakeholders (e.g., students or teachers), still pose challenges in this emergent field (Di Mitri et al., 2018; Sharma et al., 2019).

CrossMMLA provides participants with hands-on experience in gathering data from learning situations using wearable apparatuses (e.g., eye-tracking glasses, wristbands), non-invasive devices (e.g., cameras) and other technologies (in the morning half of the workshop). In addition, we will demonstrate how to analyze/annotate such data, and how machine learning algorithms can help us to obtain insights about the learning experience (in the afternoon half). CrossMMLA provides opportunities, not only to learn about exciting new technologies and methods, but also to share participants' own practices for MMLA, and meet and collaborate with other researchers in this area.

2 CROSSMMLA HISTORY AND DEVELOPMENTS

CrossMMLA continues a recently-established, but already very consistent tradition of workshops on MMLA and CrossLAK, organized at both EC-TEL and LAK conferences. These past events have leveraged a variety of formats, from hands-on learning experiences and tutorials, based on participant contributions/papers, as well as conceptual and community-building activities (which have eventually led to the creation of a Special Interest Group within Society of Learning Analytics Research - SOLAR CrossMMLA SIG¹).

The CrossMMLA community aims to become the focal point of contributions coming from a variety of fields (e.g., learning, HCI, data science, ubiquitous computing). Prior to the CrossMMLA event, we launch a call for submissions that shapes the hands-on activities to be performed. The contributions normally belong in one or more of the following categories:

1. Data gathering setups and prototypes (e.g., the use of the Multimodal Learning Hub and EEGlass).
2. Data analysis/annotation methods and tools (e.g., Visual Inspection Tool, coding schemas and “grey-box” analyses).
3. Learning activities/Pedagogical designs that could benefit from CrossMMLA techniques.
4. Examples of CrossMMLA research designs or case studies.

¹ Multimodal Learning Analytics Across Spaces Special Interest Group (SOLAR CrossMMLA SIG): <https://www.solaresearch.org/community/sigs/crossmmla-sig/>

During the CrossMMLA events, there is a formation of teams that then engage in different CrossMMLA projects. These teams use the aforementioned contributions to define learning scenarios or learning activities to be performed, the research questions to be investigated through the use of CrossMMLA, and the data gathering, annotation and analysis to be undertaken during the workshop.

Announcements and future CrossMMLA calls are available here: <http://crossmmla.org/>

3 OBJECTIVES AND INTENDED OUTCOMES

It is expected that at the end of the CrossMMLA workshop, participants engage with:

- The state-of-the-art ideas, designs and implementations of CrossMMLA systems.
- Capture, analyze and report multimodal data on-the-spot.
- Contribute and shape the research agenda and future of CrossMMLA community.

Aside from the (intangible, but very important) learning of participants about CrossMMLA, and the strengthening of the SoLAR Special Interest Group on CrossMMLA, the workshop also has targeted the following two tangible outcomes:

1. Based on the contributions of the participants we provide a catalogue of shared community knowledge.
2. Based on the learning activities tested in the workshop, and the rest of the hands-on activities, an open “CrossMMLA dataset” will be made available to the community (through the SIG/Workshop website or other European open science repositories)

All contributions and materials are made available on “LAK Companion Proceedings”. Organisers are planning to create a collaborative contribution describing the consensus reached during the workshop. Based on the outcomes of the workshop and participants interest, similarly with previous versions of Cross-MMLA, we will consider proposing a special issue in an international journal (e.g., JLA, CHB, BIT or else).

REFERENCES

- Wolfe, J. M., et al., (2015). *Sensation & perception* (4th ed.). Sunderland, MA: Sinauer Associates.
- Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238.
- Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). The Big Five: Addressing Recurrent Multimodal Learning Data Challenges. In *The 8th International Learning Analytics & Knowledge Conference* (pp. 420-424). SoLAR.
- Donnelly, P. J., et al. (2016). Automatic teacher modeling from live classroom audio. *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*, 45–53. ACM.
- Echeverria, V., Martinez-Maldonado, R., & Buckingham Shum, S. (2019). Towards Collaboration Translucence: Giving Meaning to Multimodal Group Data. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (p. 39). ACM.
- Giannakos, M. N., Sharma, K., et al., (2019). Multimodal data as a means to understand the learning experience. *International Journal of Information Management*, 48, 108-119.

- Ochoa, X. (2017). Multimodal learning analytics. *Handbook of Learning Analytics*, 129–141.
- Ochoa, X. et al. (2018). The RAP System: Automatic Feedback of Oral Presentation Skills Using Multimodal Analysis and Low-cost Sensors. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*. ACM. 360-364.
- Prieto, L. P., Sharma, K., Kidzinski, Ł., Rodríguez - Triana, M. J., & Dillenbourg, P. (2018). Multimodal teaching analytics: Automated extraction of orchestration graphs from wearable sensor data. *Journal of computer assisted learning*, 34(2), 193-203.
- Scherer, S., Worsley, M., & Morency, L. P. (2012). 1st international workshop on multimodal learning analytics. *14th ACM International Conference on Multimodal Interaction, ICMI 2012*.
- Sharma, K., Papamitsiou, Z., & Giannakos, M. (2019). Building pipelines for educational data using AI and multimodal analytics: A “grey - box” approach. *British Journal of Educational Technology*, 50(6), 3004-3031.
- Spikol, D., Ruffaldi, E., Dabisias, G., & Cukurova, M. (2018). Supervised machine learning in multimodal learning analytics for estimating success in project - based learning. *Journal of Computer Assisted Learning*, 34(4), 366-377.

Context-aware Multimodal Learning Analytics Taxonomy

Maka Eradze

School of Digital Technologies, Tallinn University, Estonia

Faculty of Engineering, University of Naples Federico II, Italy

maka@tlu.ee

María Jesús Rodríguez Triana, Mart Laanpere

School of Digital Technologies, Tallinn University, Estonia

mjrt@tlu.ee, martlaa@tlu.ee

ABSTRACT: Analysis of learning interactions can happen for different purposes. As educational practices increasingly take place in hybrid settings, data from both spaces are needed. At the same time, to analyse and make sense of machine aggregated data afforded by Technology-Enhanced Learning (TEL) environments, contextual information is needed. We posit that human labelled (classroom observations) and automated observations (multimodal learning data) can enrich each other. Researchers have suggested learning design (LD) for contextualisation, the availability of which is often limited in authentic settings. This paper proposes a Context-aware MMLA Taxonomy, where we categorize systematic documentation and data collection within different research designs and scenarios, paying special attention to authentic classroom contexts. Finally, we discuss further research directions and challenges.

Keywords: multimodal learning analytics, human-labelled observations, automated observations, classroom observations, technology-enhanced classrooms, learning design, context

1 INTRODUCTION AND BACKGROUND

As teaching and learning processes most often take place blended learning settings, to create a holistic picture of educational context and analyse these processes for different purposes, different data sources and collection methods are needed. Learning interaction (between people and/or with artefacts) has been an important part of educational research. While some decades ago, researchers focused on face-to-face interactions and used traditional data-collection techniques such as observations, technological advancements led attention to Technology-enhanced Learning (TEL) researchers towards digital interactions, as it is illustrated by the appearance of the Learning Analytics (LA) community. Thus, both research trends often cover only one part of the educational process due to the data available. The Multimodal Learning Analytics (MMLA) field emerged as a response to this need, combining different data-sources from different spaces, e.g., with the help of

sensors, EEG devices etc. At the same time, to guide the data collection and analysis process, human inference and contextual information (such as learning designs where teachers report about their intentions, actors, roles, media use and other information about the learning context) are often needed (Hernández-Leo, Rodríguez Triana, Inventado, & Mor, 2017). To address this need, previous research proposes to benefit from the synergistic LD and LA relationship, where LD contextualizes data analysis and LA informs LD.

The Learning Analytics (LA) community emerged with the widespread adoption of digital learning platforms, mainly focusing on the analysis of digital interactions (Ochoa & Worsley, 2016). However, depending on the learning activity, meaningful interactions may also not be tracked in these spaces, narrowing down the interaction analysis to the data available in the digital platforms that can lead to a street-light effect (Freedman, 2010). To respond to this limitation, a new wave of Multimodal Learning Analytics (MMLA) community promotes the collection and analysis of different data sources across spaces (Blikstein & Worsley, 2016). Typically, MMLA datasets include not only log data, but also data generated by sensors located in mobile and wearable devices (Ochoa & Worsley, 2016). To make sense of the MMLA data, input from humans is often used; human-mediated labelling is often used to relate raw data to more abstract constructs (Worsley et al., 2016)(Di Mitri, Schneider, Klemke, Specht, & Drachsler, 2019). At the same time, analytics approaches need theory (Joksimović, Kovanović, & Dawson, 2019) to create a hypothesis space (Di Mitri, Schneider, Specht, & Drachsler, 2018). Moreover, contextual information such as the learning design can guide the data collection and interpretation (Lockyer & Dawson, 2011)(Rodríguez-Triana, Martínez-Monés, Asensio-Pérez, & Dimitriadis, 2013). However, it is worth noting that in authentic settings LD may not be available due to different limitations and adoption problems (Dagnino, Dimitriadis, Pozzi, Asensio-Pérez, & Rubia-Avi, 2018)(Lockyer, Heathcote, & Dawson, 2013; Mangaroska & Giannakos, 2018).

Traditional human-mediated data collection methods, such as observations can also respond to the aforementioned need for contextual information, as they are inherently highly contextual. Through observational methods, quantitative and qualitative data can be systematically collected and analysed (Cohen, Manion, & Morrison, 2018)(Eradze, Rodríguez Triana, & Laanpere, 2017). However, despite the richness of observational data, several constraints prevent researchers and practitioners from applying them (e.g., time-consuming data collection and analysis, intrusive approach, difficulties registering fine-grain events or multiple events at the same time, etc). Therefore, educational research and practice may benefit from aligning traditional (human-labelled) and modern (automated) classroom observations; thanks to the evidence collected from the physical space, they can support the triangulation, contextualization and sensemaking of MMLA data. On the one hand, observations could aid the MMLA contextual and methodological needs, and on the other MMLA could alleviate the complexity and workload of human-driven observations: enrich the data, speed up the observation process by automatization or gather evidence on indicators unobservable to the human eye, as already indicated by previous authors (Anguera, Portell, Chacón-Moscoso, & Sanduvete-Chaves, 2018)(Bryant et al., 2017). Furthermore, technological solutions may further reinforce the use of specific coding schemas, contributing to the quality and availability of the data; speed up the process of observations (Kahng & Iwata, 1998), and enhance validity and reliability of data (Ocumpaugh et al., 2015).

Based on the overviewed community challenges and concerns rooted in previous research, to provide a holistic picture on teaching and learning processes and with a systematic picture on the use of MMLA in different scenarios, this research has connected two research paradigms (traditional and modern) based on systematic, human-labelled and automated observations. More concretely, we explore synergies between these two approaches in authentic, blended, TEL classroom settings. Also, to reinforce the contextualization, whenever available, we propose to use the LD, reflecting the pedagogical grounding and the teacher intentions leading to that activity. Connecting these three factors: human-mediated, automated observations and LD contextualization is not a trivial task, and special attention needs to be paid to the specificities, meaning, affordances, constraints and quality of the data sources, as well as LD availability challenges.

To envision the data collection and documentation process, we propose a *Context-Aware MMLA Taxonomy*. The presented taxonomy classifies different research designs depending on how systematic the documentation of the learning design and the data collection have been. The following section will overview the taxonomy and the final chapter of the paper will close with a discussion detailing further research directions and challenges.

2 CONTEXT-AWARE MULTIMODAL LEARNING ANALYTICS TAXONOMY

To provide a contextualized and holistic view of the teaching and learning activities taking place in TEL classrooms, connecting two research paradigms (Daniel, 2019), this paper proposes a *Context-aware MMLA Taxonomy* to support the alignment of LD, human and automated observations (MMLA). In this taxonomy, in line with previous research indicating to LA adoption challenges (Buckingham Shum, Ferguson, & Martinez-Maldonado, 2019), we regard authentic learning contexts as a baseline, anchoring scenario. The taxonomy (Figure 1) classifies human-labelled and automated data collection on two axes: systematic documentation and data-collection, viewing authentic cases as a baseline for data collection and analysis. These two axes represent context-awareness (systematic documentation) and rigorous quantitative classroom observation data collection (systematic data collection) to enable alignment of data sources and rich MMLA analysis.

Ideal - Systematic documentation and data collection: In the most desirable case, the learning design (including actors, roles, resources, activities, timeline, and learning objectives) is set up-front and documented in an authoring tool. Then, during the enactment, logs are collected automatically from the digital space and systematic observations from the physical one. During the enactment, the additional layer of enactment lesson structure is inferred through unstructured observations. To ensure the interoperability, actors and objects should be identifiable (across the learning design, logs and observations) and timestamps for each event should be registered. Once the data is aggregated in a multimodal dataset, further analysis can be executed.

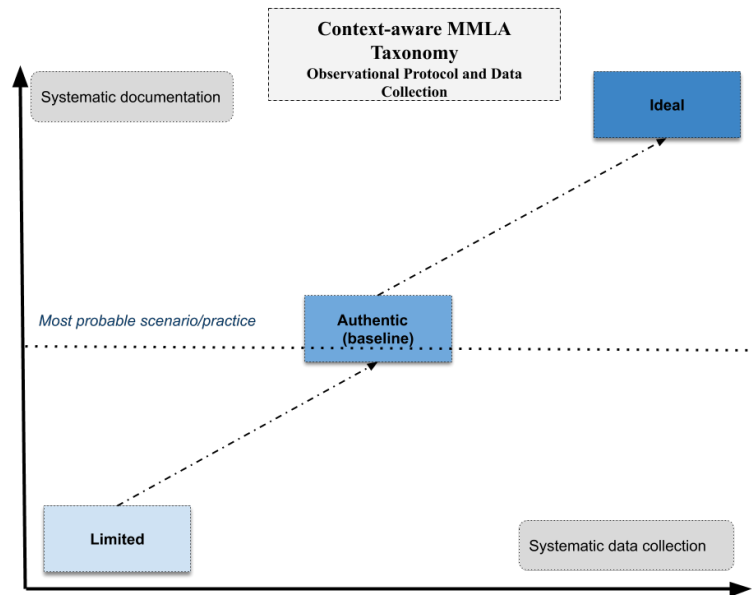


Figure 1: Context-Aware MMLA Taxonomy

Authentic (baseline) - Non-systematic documentation but systematic data collection: We regard this level as a compromise between the limitations of authentic settings but still rich in terms of data. Here, the predefined learning design cannot be automatically used to guide the analysis (either because of its format or because it is not available). However, the timestamped lesson structure is inferred by the observer. Therefore, the actors are not identifiable across observations and digital traces. Nevertheless, both structured observations and logs are systematically gathered and collected in the Learning Record Store using a common format (e.g., xAPI). These conditions will enable the application contextualized analysis on a more baseline level, using multimodal analytics.

Limited - Non-systematic documentation or data collection: Data collection happens non-systematically. As in the previous case, no information about the learning design is available (i.e., actors are not known). In terms of the design of the data collection, the protocol with corresponding codes may not be predefined, and semi-structured (non-systematic) observations are used. Thus, even if logs are systematically gathered, the lack of systematization of the observations hinder the application of multimodal data analysis. Although this is not an advisable scenario, logs and observations can be analysed independently and still provide an overview of what happened in the physical and digital planes. Besides, even if observations are done systematically, if the vocabulary (actors, objects and actions) are not agreed across datasets, then the potential of the multimodal analysis could be limited.

3 DISCUSSION, CHALLENGES AND FUTURE RESEARCH

This paper overviewed modern challenges in MMLA community underlying data contextualization and sense-making needs, especially in authentic learning scenarios. Based on these challenges and problems we suggested aligning modern and traditional data collection methods (human-labelled and automated) and LD. As researchers and practitioners need to take into account authentic learning settings in MMLA data collection, we proposed the *Context-aware Multimodal Taxonomy* to classify different levels of data collection and documentation, for different research designs. It is

worth noting that we also created specific conceptual and technological tools (Eradze & Laanpere, 2017; Eradze, Rodríguez-Triana, & Laanpere, 2017). Both, the taxonomy and tools have been evaluated in authentic settings (corresponding to the *baseline scenario*) through an iterative analysis of multimodal data (human-labelled and automated observations) involving different qualitative sources such as teacher reflections and qualitative observations. Preliminary results show that, in authentic settings, the baseline scenario was useful for two-level contextualization: observed lesson structure, human-labelled observations. At the same time, in this specific case, systematic human-labelled observations introduced additional semantics, pedagogical constructs, and indicate to the potential of using theoretical constructs in the automated observation data-sets through (validated) coding schemas. This factor further contributes to the creation of hypothesis space.

However, to enable alignment of MMLA observations and LD, in *ideal scenarios* (see Figure 1) and to facilitate the adoption of MMLA in the context of classroom observations by final users, there is a need for further reinforcement for sense-making and analysis to enable actionable insights based on MMLA data. To reach that goal, it would be necessary to create MMLA architectures and pipelines to integrate MMLA data and visualize it in a dashboard. In this regard, the on-going MMLA research efforts (Schneider, Di Mitri, Limbu, & Drachsler, 2018; Shankar et al., 2019) look very promising. At the same time, further research is needed for the pedagogically-grounded and theory-driven analysis of data and understanding how the *Context-aware MMLA taxonomy* and the related solutions can inform the teaching practice.

REFERENCES

- Alison Bryant, J., Liebeskind, K., & Gestin, R. (2017). Observational Methods. In *The International Encyclopedia of Communication Research Methods* (pp. 1–10). <https://doi.org/10.1002/9781118901731.iecrm0171>
- Anguera, M. T., Portell, M., Chacón-Moscoso, S., & Sanduvete-Chaves, S. (2018). Indirect observation in everyday contexts: Concepts and methodological guidelines within a mixed methods framework. *Frontiers in Psychology*, 9(JAN), 13. <https://doi.org/10.3389/fpsyg.2018.00013>
- Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining : using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220–238. <https://doi.org/http://dx.doi.org/10.18608/jla.2016.32.11>
- Buckingham Shum, S., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-Centred Learning Analytics. *Journal of Learning Analytics*, 6(2), 1–9. <https://doi.org/10.18608/jla.2019.62.1>
- Cohen, L., Manion, L., & Morrison, K. (2018). Research methods in education, 8th ed. In *Routledge*. <https://doi.org/10.1080/19415257.2011.643130>
- Dagnino, F. M., Dimitriadis, Y. A., Pozzi, F., Asensio-Pérez, J. I., & Rubia-Avi, B. (2018). Exploring teachers' needs and the existing barriers to the adoption of Learning Design methods and tools: A literature survey. *British Journal of Educational Technology*, 49(6), 998–1013. <https://doi.org/10.1111/bjet.12695>
- Daniel, B. K. (2019). Big Data and data science: A critical review of issues for educational research. *British Journal of Educational Technology*. <https://doi.org/10.1111/bjet.12595>
- Di Mitri, D., Schneider, J., Klemke, R., Specht, M., & Drachsler, H. (2019). Read Between the Lines: An Annotation Tool for Multimodal Data for Learning. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 51–60. ACM.
- Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning*, 34(4), 338–349. <https://doi.org/10.1111/jcal.12288>

- Eradze, M., & Laanpere, M. (2017). Lesson observation data in learning analytics datasets: Observata. In: Lavoué É., Drachsler H., Verbert K., Broisin J., Pérez-Sanagustín M. (Eds) *Data Driven Approaches in Digital Education. EC-TEL 2017. Lecture Notes in Computer Science, Vol 10474. Springer, Cham, 10474 LNCS*, 504–508. https://doi.org/10.1007/978-3-319-66610-5_50
- Eradze, M., Rodríguez-Triana, M. J., & Laanpere, M. (2017). Semantically Annotated Lesson Observation Data in Learning Analytics Datasets: a Reference Model. *Interaction Design and Architecture(s) Journal*, 33(75–91), 75–91. Retrieved from http://www.mifav.uniroma2.it/inevent/events/idea2010/doc/33_4.pdf
- Eradze, M., Rodríguez Triana, Jesús, M., & Laanpere, M. (2017). How to aggregate lesson observation data into learning analytics dataset? *Joint Proceedings of the 6th Multimodal Learning Analytics (MMLA) Workshop and the 2nd Cross-LAK Workshop Co-Located with 7th International Learning Analytics and Knowledge Conference (LAK 2017)*. Vol. 1828. No. CONF. CEUR, 2017., 1828, 74–81. CEUR.
- Freedman, D. H. (2010). Why scientific studies are so often wrong: The streetlight effect. *Discover Magazine*, 26.
- Hernández-Leo, D., Rodríguez Triana, M. J., Inventado, P. S., & Mor, Y. (2017). Preface: Connecting Learning Design and Learning Analytics. *Interaction Design and Architecture(s) Journal*, 33(Ld), 3–8. Retrieved from <https://infoscience.epfl.ch/record/231720>
- Joksimović, S., Kovanović, V., & Dawson, S. (2019). The Journey of Learning Analytics. *HERDSA Review of Higher Education*, 6, 27–63.
- Kahng, S., & Iwata, B. A. (1998). Computerized systems for collecting real-time observational data. *Journal of Applied Behavior Analysis*, 31(2), 253–261.
- Lockyer, L., & Dawson, S. (2011). Learning designs and learning analytics. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge - LAK '11*, 153. <https://doi.org/10.1145/2090116.2090140>
- Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist*, 57(10), 1439–1459.
- Mangaroska, K., & Giannakos, M. N. (2018). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies*, 1–1. <https://doi.org/10.1109/TLT.2018.2868673>
- Ochoa, X., & Worsley, M. (2016). Augmenting Learning Analytics with Multimodal Sensory Data. *Journal of Learning Analytics*, 3(2), 213–219.
- Ocuppaugh, J., Baker, R. S., Rodrigo, M. M., Salvi, A., van Velsen, M., Aghababayan, A., & Martin, T. (2015). HART. *Proceedings of the 33rd Annual International Conference on the Design of Communication - SIGDOC '15*, 1–6. <https://doi.org/10.1145/2775441.2775480>
- Rodríguez-Triana, M. J., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2013). Towards a script-aware monitoring process of computer-supported collaborative learning scenarios. *International Journal of Technology Enhanced Learning*, 5(2), 151–167. <https://doi.org/10.1504/IJTEL.2013.059082>
- Schneider, J., Di Mitri, D., Limbu, B., & Drachsler, H. (2018). Multimodal learning hub: A tool for capturing customizable multimodal learning experiences. *European Conference on Technology Enhanced Learning*, 45–58. Springer.
- Shankar, S. K., Ruiz-Calleja, A., Serrano-Iglesias, S., Ortega-Arranz, A., Topali, P., & Martínez-Monés, A. (2019). *A Data Value Chain to Model the Processing of Multimodal Evidence in Authentic Learning Scenarios*.
- Worsley, M., Abrahamson, D., Blikstein, P., Grover, S., Schneider, B., & Tissenbaum, M. (2016). Situating multimodal learning analytics. *12th International Conference of the Learning Sciences: Transforming Learning, Empowering Learners, ICLS 2016*, 1346–1349. International Society of the Learning Sciences (ISLS).

Towards Teacher Orchestration Load-aware Teacher-facing Dashboards

Ishari Amarasinghe, Milica Vujovic and Davinia Hernández-Leo

TIDE, ICT Department, Universitat Pompeu Fabra, Barcelona

{ishari. amarasinghe, milica. vujovic, davinia.hernandez-leo}@upf.edu

ABSTRACT: In this workshop paper, we report a study conducted to investigate the use of tracking technologies to measure the teachers' orchestration load when conducting co-located collaborative learning activities. We distinguish the orchestration load experienced by the teachers in the absence and presence of teacher supporting tools, i.e. teacher-facing dashboards. Electrodermal activity (EDA) sensor and other multimodal data including observations, log data and subjective responses to questionnaires have been collected to measure the teachers' orchestration load in authentic collaborative learning scenarios. This workshop paper presents the study context, quantitative and qualitative data collection process undertaken and other considerations in detail.

Keywords: Computer-Supported Collaborative Learning, orchestration load, dashboards, MMLA, electrodermal activity (EDA).

1 INTRODUCTION

In the domain of Technology-Enhanced Learning (TEL) the notion of orchestration refers to “how a teacher manages, in real-time multi-layered activities in a multi-constraint context” (Dillenbourg, 2013). In the context of Computer-Supported Collaborative Learning (CSCL), orchestrating collaboration is an essential yet a challenging task which demands teachers' continuous monitoring, guidance and interventions across different social levels (e.g., individual, group and class level). On the other hand, the application of Learning Analytics (LA) tools in the context of CSCL has currently gained heightened attention (Jivet, Scheffel, Specht & Drachsler, 2018). By capturing, analyzing and visualizing data traces that represent students' collaborative interactions in real-time, LA offers the possibility for teachers to obtain a deeper understanding of the process of collaboration and student activity engagement (Jivet et al., 2018). Towards this end, teacher-facing dashboards have been deployed within CSCL environments as a supporting tool with objectives of building awareness and facilitating teachers' productive intervention towards groups that require immediate attention (van Leeuwen, 2015).

However, the number of studies that investigate whether the addition of teacher-facing dashboard applications influence orchestration load of the teacher is scarce. It is essential to study how the addition of such supporting tools contribute to the orchestration load of the teachers, as it will facilitate to elicit useful design guidelines that can guide the development of teacher support tools that may help reduce the orchestration load experienced. Towards this end, this workshop paper presents details of an experiment conducted to study how data collected in different modalities can be used as indicators to measure teachers' orchestration load in co-located CSCL settings.

2 STUDY DESIGN

2.1 Participants

Two female teachers from a Spanish University participated in the experiments. Teachers had prior experience in conducting collaborative learning activities and have used dashboard applications to orchestrate collaboration. Each teacher conducted three collaborative learning activities and students from the respective classes took part in the study with their informed consent. Each collaborative learning activity lasted around nine minutes.

2.2 Procedure

Before the classroom trials, to generate appropriate baseline data, teachers were asked to wear the EDA sensor for two hours for three days and mark the events of those days that were out of the ordinary working activities. The measurement of two hours per day, was taken during working hours when teachers conduct work activities outside of the classroom. In this way workload exists, but it is not affected by the teaching itself and the presence of students and tools used during lessons.

After collecting baseline data, collaborative learning activities were conducted in classroom sessions. A web-based tool called PyramidApp (Manathunga & Hernández-Leo, 2018). that implements the Pyramid pattern based on collaborative learning activities was used to design and deploy collaboration. In the experimental condition, teachers monitored and orchestrated the group activities using a teacher-facing dashboard; whereas the dashboard was not available in the control condition. The experimental condition was subdivided into two conditions based on the presence of certain warnings in the dashboard. For instance, in Dashboard condition I, the dashboard generated several warnings when; 1) students answers does not contain any keyword that was stated by the teacher during activity design time, 2) students skipped answer submissions, 3) students require more time for collaboration, 4) collaborative learning activity reaches the end. In the Dashboard condition II, the aforementioned warnings were turned off, but all the other features of the dashboard were available.

2.3 Data collection and analysis

At the beginning of each collaborative learning session we attached the Shimmer3 GSR+ sensor to the teacher by connecting two electrodes to the wrist and putting arm band that holds the sensor around the teacher's arm. The sensor is placed on the non-dominant hand to avoid discomfort to the teacher and reduce the noise produced by the movement (see Figure 1).

The sensor is mounted before the beginning of the activity and removed right after. Recording begins as soon as the sensor is removed from the docking station connected to the computer, so that the signal captured between this moment and the beginning of the activity, is being removed from the analysis. The same action is applied at the end of the recording. Signal captured between the end of the activity and connecting the sensor back to the docking station (end of recording) is being removed. Data transfer from the device was conducted immediately after the activity. Moreover, teacher's behaviour during every session was recorded either using a video camera or by a researcher taking observation notes based on the unique requirements of each classroom session. In the experimental sessions teacher's dashboard actions were automatically logged. Teachers'

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

subjective measurements of the cognitive load experienced in both control and experimental sessions were also collected using NASA's TLX questionnaire (Hart & Staveland, 1988). Stimulated-recall interviews were also conducted with the teacher to better understand their orchestration requirements and pedagogical decision-making (see Figure 2).

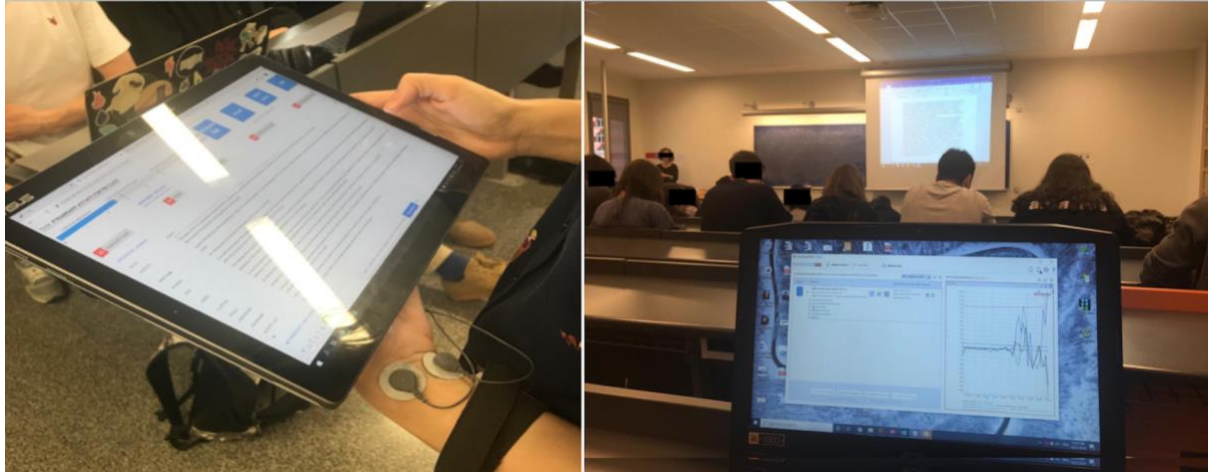


Figure 1: A teacher wearing the Shimmer3 GSR+ sensor during a classroom session (left) and data collection in a co-located collaborative learning setting (right)

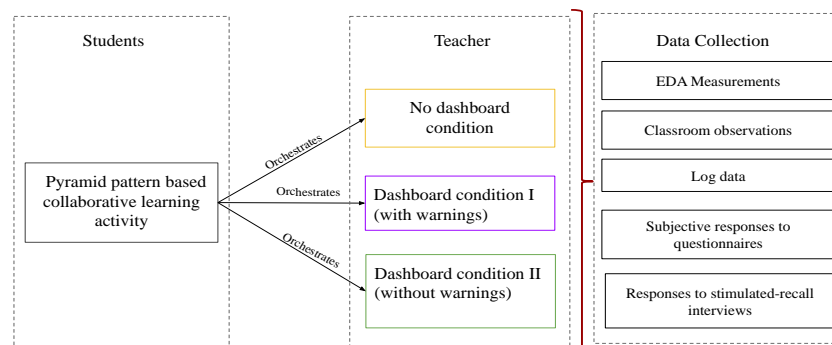


Figure 2: Different experimental conditions and data collection

3 CONCLUSIONS & FUTURE WORK

The addition of supporting tools to synchronous collaborative settings could facilitate teachers to diagnose collaboration (van Leeuwen, 2015). LA dashboards have been seen as a promising tool that can assist to raise teacher awareness, reflection and sense-making on peer learning activity engagement and to impact behavior (van Leeuwen, 2015). In this study we have collected qualitative and quantitative data in different modalities in order to measure orchestration load experienced by the teachers. A mixed-method approach will be used with the triangulation of quantitative and qualitative data to warrant results about the three conditions. We will analyse the collected data to explore how multimodal data can be used as indicators to measure teachers' orchestration load in order to propose orchestration load aware design guidelines for teacher-facing dashboards.

ACKNOWLEDGEMENTS

This work has been partially funded by FEDER, the national research agency of the Spanish Ministry of Science, Innovations and Universities MDM-2015-0502, TIN2017-85179-C3-3-R.

REFERENCES

- Dillenbourg, P., Nussbaum, M., Dimitriadis, Y., & Roschelle, J. (2013). Design for classroom orchestration. *Computers & Education*, 69(0), 485-492.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52, (pp. 139-183).
- Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018). License to evaluate: Preparing learning analytics dashboards for educational practice. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 31-40).
- Manathunga, K., & Hernández-Leo, D. (2018). Authoring and enactment of mobile pyramid-based collaborative learning activities. *British Journal of Educational Technology*, 49(2), 262-275.
- van Leeuwen, A. (2015). Learning analytics to support teachers during synchronous CSCL: Balancing between overview and overload. *Journal of learning Analytics*, 2(2), 138-162.

MMLA Approach to Track Collaborative Behavior in Face-to-Face Blended Settings

Pankaj Chejara, Reet Kasepalu, Shashi Kant Shankar, Luis P. Prieto, María Jesús Rodríguez-Triana

Tallinn University

pankajch@tlu.ee, reetkase@tlu.ee, shashik@tlu.ee, lprisan@tlu.ee, mjrt@tlu.ee

Adolfo Ruiz-Calleja

University of Valladolid

adolfo@gsic.uva.es

ABSTRACT: Collaborative learning is a complex and multifaceted phenomenon which requires teachers to pay close attention to their students in order to understand the underlying learning process and to offer needed help. However, in authentic settings with multiple groups, it becomes extremely difficult for teachers to observe each group. This paper presents our current MMLA prototype, which allows the collection, analysis and visualization of two types of data from students: audio and logs. We showcase our idea using a Raspberry Pi-based prototype (named CoTrack) for capturing and understanding the students' behavior during face-to-face blended collaborative learning situations. More specifically, CoTrack captures audio data together with software logs captured from their activities using a digital tool Etherpad. Later on, the collected data collected is analyzed to extract the participation behavior across physical and digital spaces. CoTrack has been used in 2 lab and 2 authentic case studies. Preliminary results show that despite of manual set-up and accuracy problems which may emerge, practitioners have shown interest in using it in their (authentic) classroom practice.

Keywords: Multimodal Learning Analytics, Collocated Collaboration

1 INTRODUCTION

Multimodal Learning Analytics (MMLA) has offered a new perspective for understanding learning by utilizing a wide range of sensors and machine learning algorithms (Ochoa, 2017). In addition to informing how learning takes place in real-world settings, MMLA can also “generate distinctive insights into what happens when students create unique solution path to problems, interact with peers, and act in both the physical and digital space” (Blikstein & Worsley, 2016). Researchers have demonstrated the usefulness of MMLA in understanding a range of learning constructs, e.g., emotion, attention, level of expertise, collaboration behavior, and cognition (Di Mitri, Schneider, Specht, & Drachsler, 2018). However, the deployment of MMLA in authentic learning settings is extremely difficult due to the challenges of multimodal data collection and analysis (such as the complex technological set-up, multimodal data fusion, or noisy data) (Chua, Dauwels, & Tan, 2019). These issues need to be addressed in order to raise MMLA adoption in authentic settings.

Collaborative learning in face-to-face (F2F) blended settings includes usage of digital collaboration tools with F2F interactions. However, researchers have either focused on F2F or digital interactions to understand collaboration behavior, but not much work has addressed these two spaces together

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

(Rodríguez-Triana et al., 2017). Thus, showing the participation behavior in digital and physical spaces could potentially be helpful for practitioners and researchers to understand collaboration among students. However, the individual analysis of the spaces poses certain limitations: since digital contributions from the students are not taken into account in F2F interaction analyses and, viceversa, log-based LA tools miss F2F interactions. Such lack of joint analyses is in part justified by the multimodal data collection and analysis challenges (e.g. synchronization and fusion).

This paper presents an MMLA prototype -CoTrack- for data collection, data analysis, and visualization to understand collaborative learning in F2F blended settings. Concretely, CoTrack captures students' interactions from F2F discussions and written tasks through audio data and logs, respectively. Once collected, interactions from both spaces are mapped to the corresponding students in order to measure their participation (e.g., speaking time and number of edits). Finally, different visualizations are generated to enable post-hoc reflection by practitioners.

2 DATA COLLECTION

Researchers consider talk the most important resource in collaboration (Roschelle & Teasley, 1995). In fact, audio features e.g. verbal (speech) and non-verbal (pitch, energy) features are good predictors of collaboration quality and success (Praharaj, Scheffel, Drachsler, & Specht, 2018). Also, when collaboration takes place through digital means, user interactions have been extensively used to understand collaboration behavior. These findings led us to capture both audio data and digital traces. The another rationale for restricting the prototype to audio and logs is to make prototype and its deployment simple and cheaper (as its target is eventually wide authentic settings deployment).

Our idea for the research prototype is motivated by the work of (Noel et al., 2018) which explored the collaboration behavior during collaborative writing activities. This work used the Raspberry Pi module with Microphone array to capture audio data during collaborative writing. Their work focused on F2F interactions by capturing audio data during the collaborative writing and generated visualization (e.g. social network). Our prototype, however, considers digital logs as well collected from students' writing activities in Etherpad (collaborative tool). We developed a similar prototype using Raspberry Pi¹ (3 Model B+) (Learning, 2016) and 4-Mic Microphone array (ReSpeaker) to capture social interaction pattern through audio data. In addition, we developed a plugin to collect students' interactions in a real-time collaborative editor tool: Etherpad. Figure 1 offers an overview of our prototype for capturing the multimodal data during collaboration activity.

For preprocessing, CoTrack uses VAD (Voice Activity Detection) and DoA (Direction of Arrival) algorithms (from library shipped with microphone) to associate the captured audio data with the corresponding student (as each student sits at a particular degree around CoTrack). This data contains direction (in degrees) from which the sound is detected for every 200 ms duration. Each student is represented by an alias name (e.g., user-1, user-2, user-3, user-4). To map the Etherpad logs to students, we collect IP addresses before the collaboration activity. Later, CoTrack extracts features such as speaking time and sequence of who spoke after whom. These measures are computed for different

¹ <https://www.raspberrypi.org/products/raspberry-pi-1-model-b-plus/>

time windows (e.g. 2 min, 5 min, 15 min). From Etherpad logs, two features are extracted by the current version of the prototype: number of characters added and number of characters deleted. In addition, features from Etherpad logs (e.g. number of chars added or deleted, text), are merged with audio features (e.g. speaking time, speaking turns). These extracted features are then stored in a database for the purpose of analysis.

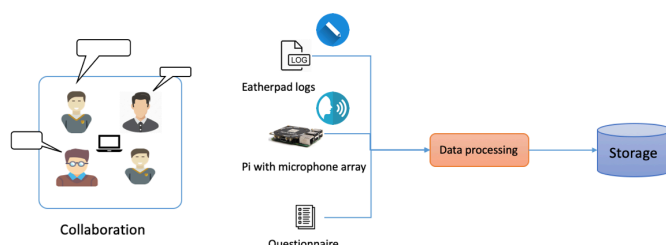


Figure 1: Data collection using CoTrack

3 DATA ANALYSIS

In the analysis phase, we perform an exploratory analysis addresses the following research questions:

- i. Which features from collected data are good predictors of collaboration?
- ii. How the collected multimodal data can be used to understand the participation behavior?
- iii. How useful are the generated multimodal data visualizations for understanding participation behavior?

To address these research questions, we developed an algorithm to process and visualize the activity traces and conducted a semi-structured focus-group interview with the teachers.

4 DATA VISUALIZATION

In this phase, we visualize the participation behavior captured through audio data and digital traces (Etherpad logs). Particularly, the current version of the prototype shows the overall speaking time, speaking time for different time window, interaction network, and number of characters added or deleted in Etherpad tool for each student. Figure 2.a shows the visualization for speaking time for the time window of 5 mins. This visualization is generated from the datasets collected from one of the experiments reported in Table 1. Each student is shown with a different color.

Figure 2.b shows the overall group interaction network with their Etherpad activities. This network is generated from the speaking sequence which is basically a sequence of "who spoke after whom". Each student is represented by a node and the edge represents the interaction between students. The thickness of the edge shows the frequency of interaction. Additionally, the width of circle outer line represents the speaking time and percentage of characters added or deleted by each student shown by pie chart in the node (e.g. green: % chars added, red: % of chars deleted, grey: % of chars added or deleted by others). We also visualize features (e.g. number of edits) extracted from Etherpad logs for

the entire duration of the activity to offer participation behavior in digital space. An example is shown in Figure 2.c.

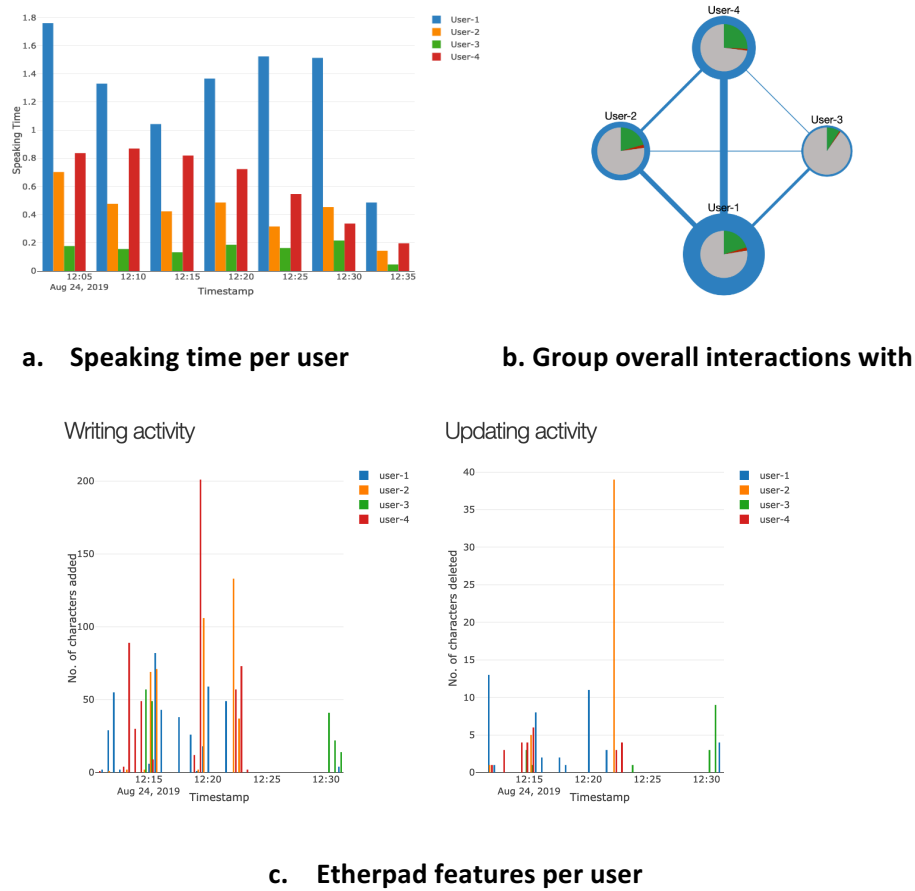


Figure 2: Visualizations

5 PRELIMINARY RESULTS

CoTrack has been used in two labs and two authentic settings. Table 1 shows the characteristics of those settings. In one of the lab settings, the practitioner herself conducted the data collection process using the web-interface of CoTrack.

Table 1: Characteristics of settings.

Settings	Group-size	Groups
Lab	4	1
Lab	4	1
Classroom	3	3
Classroom	4	2

5.1 Data Collection Protocol

In both lab and classroom settings, we first setup the CoTrack for each group and powered them up. Participants are requested to sit in a particular manner around the prototype (e.g. first participant at 45 degrees, second at 135 degrees, and so on). Then, the server machine is synchronized with the NTP server (running on one of the Pi). Once the technical infrastructure is ready, we provided a brief introduction about the prototype and the purpose of data collection. After getting the written consent for data collection from the participants, we started the Etherpad server and provided the instructions to access it on their laptops. In the classroom settings, due to time constraints, we setup the Etherpad access on each laptop before the activity. We collected the IP addresses of each laptop to map it to the corresponding participants. Once everyone had access to Etherpad, we started the audio recordings using CoTrack. For the ground truth purpose, we also video recorded the sessions. Once the activity was finished or the teacher notified about the end of activity, we stopped the audio and video recordings. Finally, we generated visualizations of collected data and showed it to the participants/teacher after the activity.

5.2 Initial Results

The current version of CoTrack utilizes only DoA data to compute speaking behavior, hence, our first aim was to investigate the feasibility of DoA. We manually annotated one group's (from authentic setting with group-size four) audio recording with speaker label, and compared it with CoTrack's results. For this comparison, we only considered annotation frames where only one participant was speaking because the CoTrack can not detect overlapping speaking activity. We determined accuracy by computing the percentage of frames (at the level of 200ms) correctly detected by CoTrack for each participant. The overall accuracy was 48%. We manually checked the video recordings to find out the reason of the low accuracy. We found that frequent moving of participant-3 towards participant-2 during the activity caused the issue of wrongly detecting audio from participant-3 as coming from participant-2. Finally, we found that sitting arrangement and movement of participants can influence the accuracy measure. Additionally, audio noise can also degrade the quality of collected DoA data.

6 CONCLUSION AND FUTURE WORK

In this paper, we presented CoTrack, an MMLA prototype for data collection, analysis and visualization of F2F collaborative learning activities. During the workshop, participants will be able to try the prototype, discuss about its pros, cons, and potential improvements, as well as learn how it could be adapted to their own CrossMMLA contexts. It will also help participants to see its benefit in understanding the social aspect of collaboration with automated data collection and analysis. In future stage of this research, we plan to use questionnaire data and collaboration quality rating schemes for collaboration measure to identify types of collaboration patterns and corresponding multimodal features.

REFERENCES

- Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220-238.

- Chua, Y. H. V., Dauwels, J., & Tan, S. C. (2019, March). Technologies for automated analysis of co-located, real-life, physical learning spaces: Where are we now?. In Proceedings of the 9th International Conference on Learning Analytics & Knowledge (pp. 11-20). ACM.
- Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning*, 34(4), 338-349. <https://doi.org/10.1111/jcal.12288>
- Learning, U. (2016). *Raspberry Pi 3: Get Started with Raspberry Pi 3 a Simple Guide TO Understanding and Programming Raspberry Pi 3 (Raspberry Pi 3 User Guide, Python Programming, Mathematica Programming)*. CreateSpace Independent Publishing Platform.
- Noel, R., Riquelme, F., Mac Lean, R., Merino, E., Cechinel, C., Barcelos, T. S., ... & Munoz, R. (2018). Exploring Collaborative Writing of User Stories With Multimodal Learning Analytics: A Case Study on a Software Engineering Course. *IEEE Access*, 6, 67783-67798. <https://doi.org/10.1109/ACCESS.2018.2876801>
- Ochoa, X. (2017). *Multimodal Learning Analytics*. Handbook of Learning Analytics, 129-141. <https://doi.org/10.18608/hla17.011>
- Praharaj, S., Scheffel, M., Drachsler, H., & Specht, M. (2018, September). Multimodal analytics for real-time feedback in co-located collaboration. In European Conference on Technology Enhanced Learning (pp. 187-201). Springer, Cham.
- Rodriguez Triana, M. J., Prieto Santos, L. P., Vozniuk, A., Shirvani Boroujeni, M., Schwendimann, B. A., Holzer, A. C., & Gillet, D. (2017). Monitoring, awareness and reflection in blended technology enhanced learning: a systematic review. *International Journal of Technology Enhanced Learning*, 9(ARTICLE), 126-150.
- Roschelle, J., & Teasley, S. D. (1995). The construction of shared knowledge in collaborative problem solving. In *Computer supported collaborative learning* (pp. 69-97). Springer, Berlin, Heidelberg.

Physiology-Aware Learning Analytics Using Pedagogical Agents

Melanie Bleck, Nguyen-Thinh Le & Niels Pinkwart

Computer Science Department

Humboldt-Universität zu Berlin

Germany

mail@melanie-bleck.de, nguyen-thinh.le@hu-berlin.de, niels.pinkwart@hu-berlin.de

ABSTRACT: Learning analytics applications consider not only the cognitive dimension, but also the physiological dimension of the learner. This paper describes a learning analytics approach that focuses on alerting the critical stress level of the learner using a pedagogical agent. For that purpose, an existing pedagogical agent was expanded by a software component, which analyses heart rate variability data to determine the cognitive load of a user and to offer support with stress reduction. The evaluation study with the physiology-aware pedagogical agent showed an improvement of learning and a reduction of stress.

Keywords: Physiological computing, Heart rate variability, pedagogical agent

1 INTRODUCTION

One of the factors that may affect learning performance is stress (Li et al., 2017). Thus, detecting and measuring stress that occurs while learning could be used to enhance learning analytics applications. In addition to proposals to different observation techniques, e.g., facial detection and video monitoring (D’Mello, 2017). Giannacos and colleagues (Giannacos et al., 2020) suggest that physiological parameters Heart Rate, blood pressure, temperature, and electrodermal activity (EDA) level can be used as a proxy to estimate learning performance. This paper focuses on specific psychological state “stress” that might have impact on the learning process. The monitoring of physiological parameters like Heart Rate Variability (HRV) is considered a relevant indicator for the stress detection (Zangroniz et al., 2018). However, handling physiological data, to what extent they can be used to analyze excessive cognitive demands and how it can be utilized in a learning analytics context are still a research gap. The research question to be investigated in this paper is how HRV data can be used by pedagogical agents to determine the stress level of the learner and to alert the learner in a learning situation.

2 METHODOLOGY

In order to investigate the specified research question, the functionality of the web-based pedagogical agent LIZA (Le & Wartschinski, 2018) aimed at improving the decision making and reasoning of the user, was extended through three different parts. The first component provides a solution to generate, save and process the HRV data, the second one analyses the data regarding stress and the third one adapts the learning situation through selected stress reduction strategies. To determine the effectiveness and benefits of the approach, the adjusted pedagogical agent was evaluated.

To use HRV parameters to determine stress level, the generation of data has to be ensured. A technical solution was provided by the wristband E4 of Empatica. Such a device was chosen to minimize the complexity of the handling and to provide a most comfortable position and positioning of the sensors (Gjoreski & Gjoreski, 2017) to reduce entry barriers while learning. Furthermore, Empatica provides a Software Development Kit to access the data via Bluetooth. Thus, the integrated photoplethmography sensor was utilized to determine the heart rate and calculate the time interval between two consecutive heartbeats (NN Interval) (Empatica Inc., 2016). These values are retrieved by a mobile application, which is also provided by Empatica and in which the functionality to transmit the current NN Interval and a timestamp to a server via HTTP-Post-Request was added. This implementation solution was necessary because of restrictions regarding data retrieval through web applications. The server is responsible for the storage of the values in a database and the processing of the NN Intervals. Is a specific time interval requested by the pedagogical agent, the suitable NN Intervals will be selected on the basis of the time stamp and the root mean square of successive differences in the heart rate (RMSSD) of these values will be determined.



Figure 1: Example for six NN intervals

$$RMSSD = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n-1} (NN_{i+1} - NN_i)^2}$$

RMSSD¹ was chosen as a metric for HRV because of the recommendation as an indicator for cognitive load in short term measurements (less than 5 minutes) by AWMF (Sammito, et al., 2014). As mentioned before, it was necessary to alter the pedagogical intervention process of the original pedagogical agent LIZA for analyzing the RMSSD data accordingly (Figure 2).

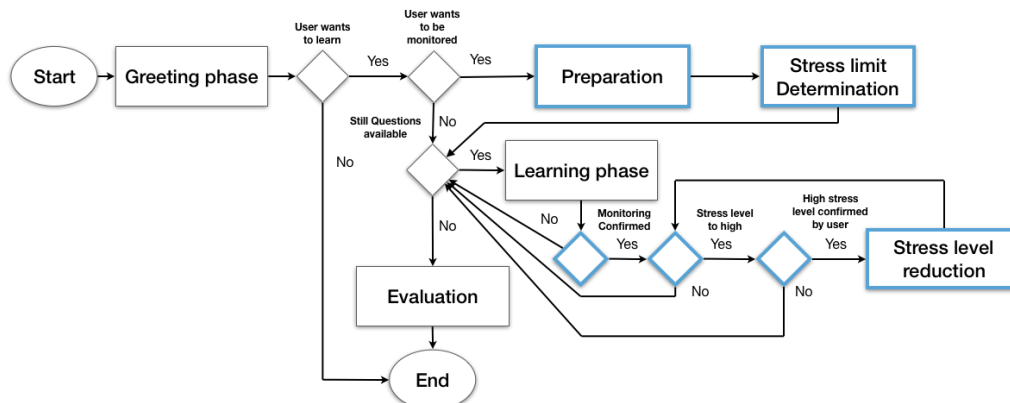


Figure 2: Intervention process

¹ NN = NN-Interval, difference in time of two consecutive R spikes; n = number of R spikes.

First, an enquire for the declaration of consent for monitoring the heart rate was added to the greeting phase. This guarantees that the original functionality remains unchanged in case of a missing measuring device. After that, the user is requested to apply and activate the wristband und start the mobile transmission application.

Since there are no generally accepted threshold values to determine the degree of a mental load of a person, a series of individual measurements has to be done (Sammito, et al., 2014). But that alone does not provide enough information to automatically identify an overload during a certain task. A range of cognitive load has to be identified and a specific threshold, when learning situation will be adapted, has to be defined. Because of that, a phase was added in the pedagogical intervention process, in which two different levels of stress are induced, the RMSSDs are calculated accordingly and used as an indicator for different stress levels. The arithmetic tasks were chosen after an analysis of induction methods for cognitive load of several research papers and they have been used widely to generate moderate stress level (Schneider et al., 2003).

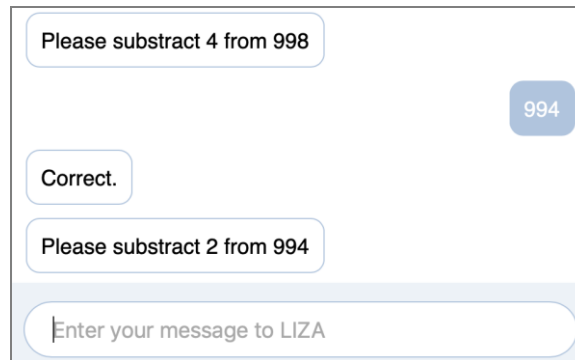


Figure 3 Arithmetic task in stress test 1

In two arithmetic stress tests (see Figure 3), with different levels of difficulty, the user had to subtract a random value from a certain number consecutively for five minutes. The result of the previous equation provides the minuend of the following. The level of difficulty is altered through the time limit for solving the equation, the number of digits of the random value and the value of the start minuend. The second and third factor determine the number of shifts during mental arithmetic, which increase the cognitive load with a growing number. Furthermore, a competition situation is created by requesting to beat LIZA in the number of correct answers under certain conditions in the second test. With that, the first test provides the RMSSD value for a moderate load, the second, which is designed with a higher difficulty level, for an overload.

Certainly, a range of cognitive load could be defined in that way, but a specific threshold, is still missing. Considering that an excessive load can result in a decrease of motivation and an abort of the learning in the long term, the learning situation has to be adapted before such a scenario materialize. Another factor, which has to be taken into account is that an adaption of the learning situation through stress reduction strategies will interrupt the process itself. So, it should be carried out as little as possible but also as much as necessary. A preliminary empirical test, where the task solutions were known and therefore low stress were induced, showed, that a threshold at 50% of the range triggers an intervention nearly every time LIZA was used. This would lead to massive interruptions of the learning process. Based on that, the threshold was increased to 2/3 of the individually defined stress range, where the intervention could be reduced to 40% of the cases.

If the current RMSSD falls below the threshold after a specific time, LIZA offers assistance in reducing the stress level through stress reduction strategies. Among different strategies (e.g., mindfulness-based stress reduction, autogenic training), two methods are proposed that are appropriate for the learning environment of a pedagogical agent. The first one distracts the user by telling jokes, the second one shows a video with relaxing content. The user decides whether it is necessary to start the offered coping process and how long the strategies are used. If the stress level is significantly reduced below the threshold, LIZA proposes the continuation of the learning process.

Figure 4 the pedagogical agent proposes two strategies for reducing stress

3 EVALUATION

The goal of the evaluation study is to determine the effectiveness and benefits of the pedagogical agent that was extended with the capability of measuring HRV and detecting critical stress level of learners. Amongst others, following hypotheses were examined: 1) The stress reduction strategies lead to the relaxation of learners; 2) The RMSSD is a suitable indicator for cognitive load; 3) The adaption of the learning process affects the learning performance. To examine these hypotheses a pre- and posttest was performed. For the study, 34 participants (10 males, 24 females) aged between 21 and 59 (mean 31 ± 11 years) were acquired and assigned to test- or control group by random. The test was conducted in a quiet environment under supervision.

Every participant was asked to use the pedagogical agent to perform two stress tests, each with a different level of difficulty, to determine the stress limit range and calculate the threshold. Then 4 tasks were given by the pedagogical agent to be solved, where every task covered a different problem of reasoning. After that, the RMSSD was calculated for the time frame of the first task block and squared with the previously determined stress limits. Only for the test group followed a stress reduction phase, if the current RMSSD fell below the threshold. Both groups continued with the posttest that required all participants to solve again 4 tasks. The reasoning problems of the pretest and posttest were the same, but the tasks were different. In the end, the participant got an evaluation of how successful the tasks were solved.

For every part of the process, the RMSSD was calculated so that the development of the indicator could be retraced. In addition, the participant had to self-assess its current state of mental load with six adjective pairs of opposite meaning after each measurement cycle. A short questionnaire for the current cognitive load (KAB) (Wagner, 2012) was used the mean of all assessments was calculated.

The first hypothesis, which covers whether stress reduction leads to the relaxation of the learner, could be partly confirmed. 90% of the participants stated in a self-assessment, which indicates a relaxation, the effect of the applied stress reduction strategies. But only in nearly 50% of the cases, the RMSSD also fell below the threshold. Possible reasons for that could be deficits in stress limit determination, insufficient choice of strategies or application time. Concerning the adequacy of the RMSSD as an indicator for cognitive load, there were rough connections between the RMSSD values and the KAB-Index, but a significant correlation between both indices could not be determined. One reason could be the error-prone self-assessment like Picard points out (Picard, 2003). Another could lie in the insufficient cognitive load, which was applied during the evaluation, to reduce the RMSSD significantly. So, the adequacy could not be confirmed unqualified and the application of other physiological parameters is suggested. Finally, the effects on learning success have to be contemplated. The test group showed a significantly higher improvement while answering the questions with comparable opportunities than the control group.

4 CONCLUSION AND OUTLOOK

This paper has demonstrated the integration of physiological factors in learning analytics using wearable sensors. It showed, that the analysis of the RMSSD to determine the cognitive load of a learner can be used to improve the learning situation. But to determine the adequacy of the RMSSD as suitable indicator further test has to be conducted. Not only shows the integrating of wearable sensors a potential to improve the learning situation, but adds also the possibility to use cognitive data beyond the use in learning analytics.

REFERENCES

- Empatica Inc. (Hrsg.). (2016). *Utilizing the PPG/BVP Signal*. Accessed 01/11/2019 <https://support.empatica.com/hc/en-us/articles/204954639-Utilizing-the-PPG-BVP-signal>
- Giannakos, M. N., Sharma, K., Papavlasopoulou, S., Pappas, I. O., Kostakos, V. (2020). Fitbit for learning: Towards capturing the learning experience using wearable sensing, *International Journal of Human-Computer Studies*, vol. 136, 102384, ISSN 1071-5819, <https://doi.org/10.1016/j.ijhcs.2019.102384>.
- Gjoreski, M., & Gjoreski, H. (2017). Monitoring Stress with a Wrist Device Using Context. *Journal of Biomedical Informatics*, 73, S. 159-170.
- Le, N.-T. & Wartschinski, L. (2018). A Cognitive Assistant for Improving Human Reasoning Skills. *International Journal of Human-Computer Studies*. 10.1016/j.ijhcs.2018.02.005.
- Li, Q., Xue, Y., et al. (2017). Analyzing and Identifying Teens Stressful Periods and Stressor Events from a Microblog. *IEEE Journal of Biomedical and Health Informatics*, 21(5), S. 1434-1448.
- Picard, R. W. (2003). Affective Computing: Challenges. *International Journal of Human-Computer Studies*, 59, S. 55-64.
- Sammito, S., Thielmann, B., et al. (2014). S2k-Leitlinie: Nutzung der Herzfrequenz und der Herzfrequenzvariabilität in der Arbeitsmedizin und Arbeitswissenschaft. *AWMF online*.

- Schneider, M., Jacobs, D. W., et al. (2003). Cardiovascular Hemodynamic Response to Repeated Mental Stress in Normotensive Subjects at Genetic Risk of Hypertension: Evidence of Enhanced Reactivity, Blunted Adaptation and Delayed Recovery. *J. of Human Hypertension*, 17(12), S. 829-840.
- Sidney D'Mello, A. K. (2017). The Affective Computing Approach to Affect Measurement. *Emotion Review*. doi:10.1177/1754073917696583
- Wagner, M. (2012). *Der Einfluss emotionaler, kognitiver und körperlicher Beanspruchung auf die Herzrate bei Jugendlichen: Eine Laborstudie*. Ph.D. Dissertation, Universität Ulm, Germany.
- Zangroniz, R., Martinez-Rodrigo, A. et al. (2018). Estimation of Mental Distress from Photoplethysmography. *Applied Science*, 8(69), S. 1-15.

Designing and implementing multimodal data collection in classroom to capture metacognition in collaborative learning

Jonna Malmberg¹, Sanna Järvelä¹, Hanna Järvenoja¹, Eetu Haataja¹, Héctor J. Pijeira-Díaz²,
Ahsen Cini¹

¹University of Oulu

² Maastricht University

Jonna.malmberg@oulu.fi, sanna.jarvela@oulu.fi, hanna.jarvenoja@oulu.fi,
eetu.haataja@oulu.fi, h.pijeiradiaz@maastrichtuniversity.nl, Ahsen.Cini@oulu.fi,
h.pijeiradiaz@maastrichtuniversity.nl

ABSTRACT: While prominent empirical research exploring the possibilities to utilize different data channels in the research of regulation in collaborative learning is emerging, we are still in the process of discovering the relevant combinations of different data sources and proper ways to combine data from different channels. This is the case particularly with metacognition. The potential of using multiple data channels lies also in their power to be transferred as a tool for providing learners ‘on the fly’ support for regulation when needed. However, an advanced understanding of the regulated learning in collaborative learning contexts, and particularly on metacognitive processes is essential to harness the benefits of technology in supporting these processes in collaborative learning.

Keywords: Metacognition, collaborative learning, multimodal data

1 INTRODUCTION

Learning processes are hard to predict or model, since learning is always situated, dependent on the learning context and the learner’s individual metacognition. Metacognitive knowledge involves learners’ perceptions of a task. It draws to prior knowledge in terms of same types of tasks and procedures needed to perform those (Winne & Hadwin, 1998). Another component of metacognition are metacognitive experiences. Metacognitive experiences constitute, for example learners’ perceptions of task difficulty. Unlike task understanding, which is thoughtful and cognitive, perception about task difficulty is reactive, and is also informative for Self-Regulation of Learning (SRL) (Winne & Hadwin, 1998). Multimodal data (e.g., physiological measures, videos, and situated self-reports) can provide a new unobtrusive way to capture learners’ metacognition without interrupting learning process (Järvelä et al., 2019). Currently, there is an accumulating evidence on how physiological measures can be used to track learning. Recent studies have shown that the level of students’ physiological arousal is related to learners’ metacognition (Hajcak, McDonald, & Simons, 2003) and achievement (Pijeira-Díaz et al. 2018). Metacognition, in turn, is related to learners’ perceptions of tasks, self and learning situations (Flavell, 1979). Yet, current research lack methods to capture the situated nature of task perceptions in the context of collaborative learning over time.

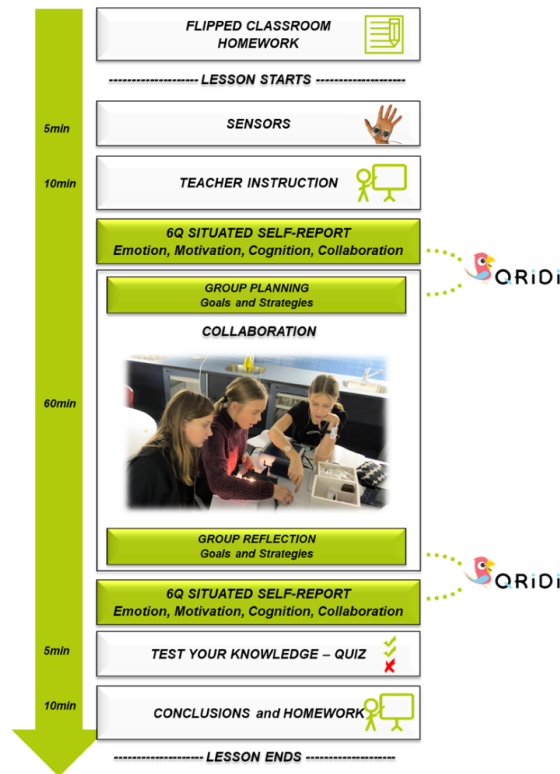
In this paper, the focus is to (1) introduce collaborative learning model designed to study processes focusing on metacognition and promoting awareness of metacognition in a secondary school science

classroom, (2) describe multimodal data collection procedure implemented in secondary school science classroom and (3) illustrate with two case examples how multimodal data has been used to capture learner's metacognition. Participants of the study were (N = 94) upper elementary school students aged 13 to 14 (58 females, 36 males) enrolled in compulsory physics course consisting altogether five lessons. In each lesson, the students collaborated in the same groups of three to four students based on the collaborative learning model. Altogether, the students had four 90 min physics lessons, once in a week and the last lesson was a collaborative exam. In addition, after each lesson, the students filled in a multiple-choice knowledge test consisting of five questions related on topics they had just learned.

2 COLLABORATIVE LEARNING MODEL

The collaborative learning model designed for science class is based on a self-regulated learning framework that provides opportunities and awareness for self-initiated regulation among individual learners and collaborative groups (authors). It utilizes technology-based environment called Qridi® (<https://kokoa.io/products/qridi>), which was designed to structure collaboration. The collaborative learning model is built on the idea of a 'flipped classroom.' Recently, the flipped classroom concept has been gaining considerable attention due to its potential to facilitate the regulation of learning (Jovanovic et al., 2019). The use of a flipped classroom in collaborative learning creates a learning setting in which students are provided opportunities to take responsibility for their own learning by familiarizing themselves with the content knowledge beforehand to prepare for collaborative learning. In the current study, the flipped classroom structure and the collaborative work were coordinated by using a Qridi® (Figure 1). However, the learning materials were not provided via Qridi®, but the students used their own regular physics books.

In the Qridi® environment, students were able to check, for example, the phase of the lesson. In our learning model, Qridi® was tailored to increase students' awareness of the collaborative learning task phases in general and, specifically, supporting their awareness of the regulation of learning. For example, Qridi® involved a 6Q tool designed to promote students' situation-specific metacognitive awareness related to task understanding and task difficulty before and after the collaborative learning. In practice, the 6Q tool consists of two 0–100 slider-scale questions where students estimate their task understanding (Schraw and Dennison, 1994), perceived task difficulty (Efklides et al. 1998).

Figure 1. Collaborative learning model

2.1 Multimodal data collection

As the students study according to collaborative learning model, multiple data sources were collected. Prior to the study, the participating students responded to trait-type questionnaires such as Metacognitive Awareness Inventory (MAI) (Schraw and Dennison, 1994) that captured their individual metacognitive beliefs. During the seven-week multichannel data collection process, students' collaborative work was followed by video recordings. Shimmer3 GSR+ sensors with 128hz sampling rate were used to measure learners' electrodermal activity (EDA) indicating arousal. The sensors were automatically synchronized with each other in the dock station before the start of each session. Students fitted the devices at the beginning of each lesson and took it off at the end. In this way, continuous EDA data was obtained for each student during the entire lesson. As one of the multiple data sources, we used the 6Q tool implemented in Qridi® to collect students' situation-specific interpretations of their metacognition in terms of task understanding and task difficulty related to each collaborative session before and after the collaborative work. Altogether the students had five physics lessons and the last one was collaborative exam.

2.2 Processing physiological data

First, files having contact issues were removed from the dataset. Second, Butterworth low pass filter with frequency 1 and order 5 was used to remove small movement artifacts from the signal. Third, Ledalab toolbox and through-to-peak analysis with minimum amplitude of 0.05μS was used for peak

detection (Benedek & Kaernbach, 2010). Number of non-specific skin conductance responses per minute (NS.SCR/min) for the session was used as a marker of arousal (Boucsein, 2012).

3 CASE EXAMPLES – WHAT ABOUT METACOGNITION?

The case examples provide insight on how to use multimodal data to investigate fluctuation of task difficulty and task understanding during collaborative learning. The first case example illustrates how individual learners' metacognitive beliefs and situation specific perceptions of task difficulty and task understanding are related on learning outcomes in the context of collaborative learning. The second case example instead focuses on exploring how individual learners' situation specific interpretations of task difficulty and task understanding are related in physiological arousal in the context of collaborative learning.

3.1 Analysis

In both of the case examples, Generalized Estimating Equations (GEE) was used. GEEs enable a general method for analyzing clustered variables and ease several assumptions of traditional regression models (Diggle, 2002; Liang & Zeger, 1998; Zeger & Liang, 1986). The GEE method does not explicitly model between-cluster variation, rather it estimates its counterpart, the within-cluster similarity of the residuals, and then uses this estimated correlation to re-estimate the regression parameters and to calculate standard error. To estimate the validity of the GEE, QIC statistics proposed by Pan (2001) allow comparisons of GEE models and selection of a correlation structure. In both case examples, normal distributions with the log link function were selected because they yielded the lowest quasi-likelihood under the independence criterion (QIC) values

3.2 How metacognitive beliefs and situated task perceptions relate for learning outcomes?

With regard to first case example, generalized estimating equations (GEE) examine the effects of individual metacognitive beliefs (MAI) and task perceptions which are task understanding (TU) and task difficulty (TD) on upper elementary school students' learning outcomes measured after each lesson.

Table 1 shows that only learners' interpretations on post-task understanding (Post TU) score can effectively predict different actualized knowledge tests that were measured after each lesson. The model fit statistics (QIC) scores was 258,986.

Table 1. GEE results model using a normal distribution with a log link function predicting students' learning outcomes

Dependent variable

Knowledge Tests

Independent variables	B (95%CI)	p value
Pre TU	0,001 (-0,001;0,003)	0,431
Post TU	0,002 (0;0,004)	0,038
Pre TD	-0,0000576 (-0,002;0,002)	0,949
Post TD	-0,001(-0,003: 0)	0,116
MAI	0,002 (-7,39E-05;0,004)	0,059

To summarize, learner individual metacognitive beliefs (which are quite static) do not predict learning outcomes, but rather learner's situation specific interpretations of the task after the collaborative learning session predicts learning outcomes at individual level.

3.3 How individuals task perceptions relate for physiological arousal when collaborative learning context is not or is considered?

With regard to second case example, generalized estimating equations (GEE) was used to examine the effects of individual understanding (TU) and task difficulty (TD) on upper elementary school students' physiological arousal (NS.SCRs in minute) during collaborative exam first at individual level (independent from the group) and second at collaborative level (exchangeable in the group).

In the light of the second case example, the results show, that when the collaborative learning context *is not considered*, task perceptions do not predict physiological arousal. The model fit statistics (QIC) scores was 10,1.

However, when the group is considered as exchangeable, the results show that learners interpretations before the task (pre-TU) score can effectively predict physiological arousal (NS.SCRs in minute) (Table 2). The model fit statistics (QIC) score was 8,943.

Table 2. GEE results model using a normal distribution with a log link function predicting students' physiological arousal

Dependent variable

NS.SCRs / minute

Independent variables	B (95%CI)	p value
Pre TU	0,006 (0,001;0,10)	0,012
Post TU	-0,004 (-0,010;0,002)	0,190
Pre TD	-0,007 (-0,016;0,001)	0,086
Post TD	-0,001(-0,10;0,008)	0,746

These two case examples shed a light in the process of discovering the relevant combinations of different data sources and proper ways to combine data to investigate metacognition. The first case example illustrates, that student characteristics, in terms of their metacognitive beliefs does not predict learning outcomes. However, the way students perceive the task after the learning situation predicts their learning outcomes.

The second example shows that when social context is taken account, task understanding predicts physiological arousal. In both examples, learner's situation specific interpretations of a task were used as an indicator of metacognition. It can be concluded, that finding (relatively) unobtrusive ways to measure and detect variations in learners task understanding as the learning proceeds, provides a fruitful venue to explore ways to implement learning analytics and to provide learners feedback and support for regulation when needed.

4 THE WORKSHOP PRESENTATION

To conclude, this presentation focuses on workshop theme: Examples of CrossMMLA research designs and case examples by presenting 1) collaborative learning model designed to capture and promote awareness of metacognition, 2) multimodal data collection implemented in science classroom and 3) providing two representative case examples of multimodal data use to capture metacognition focusing on task perceptions of learners. In the workshop, the aim is to illustrate in detail how the collaborative learning model and multimodal data collection has been designed to capture metacognition in the light of theories of regulated learning.

REFERENCES

- Boucsein, W. (2012). *Electrodermal activity* (2nd ed.). New York, NY: Springer.
- Diggle, P. J., Heagerty, P., Liang, K.-Y., & Zeger, S. L. (2002). *Longitudinal data analysis* (2nd ed.). Oxford, UK: Oxford University Press.
- Benedek, M., & Kaernbach, C. (2010). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology*, 47(4), 647–658. <https://doi.org/10.1111/j.1469-8986.2009.00972.x>
- Efklides, A., Papadaki, M., Papantoniou, G., & Kiosseoglou, G. (1998). Individual differences in feelings of difficulty: The case of school mathematics. *European Journal of Psychology of Education*, 13(2), 207–226.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, 34(10), 906.
- Ghisletta, P., & Spini, D. (2004). An introduction to generalized estimating equations and an application to assess selectivity effects in a longitudinal study on very old individuals. *Journal of Educational and Behavioral Statistics*, 29(4), 421–437.
- Hajcak, G., McDonald, N., & Simons, R. F. (2003). To err is autonomic: Error-related brain potentials, ANS activity, and post-error compensatory behavior. *Psychophysiology*, 40(6), 895–903. <https://doi.org/10.1111/1469-8986.00107>
- Jovanovic, J., Mirriahi, N., Gašević, D., Dawson, S., and Pardo, A. (2019). Predictive power of regularity of pre-class activities in a flipped classroom. *Computers & Education*, 134, 156–168.
- Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. A. (2019). What multimodal data can tell us about the students' regulation of their learning process. *Learning and Instruction*, <https://doi.org/10.1016/j>.
- Liang, K.-Y., & Zeger, S. L. (1986). Longitudinal data analysis using Generalized Linear Models. *Biometrika*, 73, 13–22.
- Pijera-Díaz, H. J., Drachsler, H., Kirschner, P. A., & Järvelä, S. (2018). Profiling sympathetic arousal in a physics course: How active are students?. *Journal of Computer Assisted Learning*, 34(4), 397–408.
- Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19(4), 460–475.
- Winne, P.H., & Hadwin, A.F. (1998). Studying as self-regulated learning. In D.J. Hacker, J. Dunlosky, & A.C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah: Lawrence Erlbaum.
- Zeger, S. L., & Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42, 121–130.

Multimodal Temporal Network Analysis to Improve Learner Support and Teaching

Mohammed Saqr

University of Eastern Finland
mohammed.saqr@uef.fi

Olga Viberg

KTH Royal Institute of Technology
oviberg@kth.se

Jalal Nouri

Stockholm University
Jalal@dsv.su.se

Solomon Oyelere

University of Eastern Finland
solomon.oyelere@uef.fi

ABSTRACT: A learning process involves interactions between learners, teachers, machines and formal and/or informal learning environments. These interactions are relational, interdependent and temporal. The emergence of rich multimodal learner data suggests the development of methods that can capture time-stamped data from multiple sources (e.g., heart rate data and eye tracking data), thus allowing researchers to examine learning as a continuous process rather than a static one. This leads us to propose a new methodological approach, the *Multimodal Temporal Network Analysis* to: i) measure temporal learner data deriving from the relevant interactions and ii) ultimately support learners and their teachers in learning and/or teaching activities.

Keywords: Multimodal learning analytics, temporal networks, social network analysis

1 INTRODUCTION

Learning occurs across both formal and informal learning settings and evolves as students interact with each other, machines, and/or with teachers, as they engage with multifaceted learning tasks. Such interactions are self- and socially regulated, temporal and interdependent (Järvelä et al., 2014). As a socially regulated process, learners' activities are facilitated or constrained by peers while they negotiate their roles, tasks and work together for the achievement of their shared goals (Malmberg, Järvelä, & Järvenoja, 2017). As a temporal process, learning follows the universal law of time, and so are the interactions and learning activities, they are forward moving, unidirectional and uniform (Saqr, Fors, & Nouri, 2019). As an interdependent process, learning activities and events are largely interdependent. To understand learning as an outcome, we need to understand the processes and sequences of past events, i.e., learning as a continuous process, which is multidimensional, complex and rich (Malmberg et al., 2017). An adequate understanding of such a process requires new innovative methods that can capture learning and its related activities as a continuous process rather than a static one. Multimodal learning analytics (MMLA) have emerged to address this issue.

1.1 Background

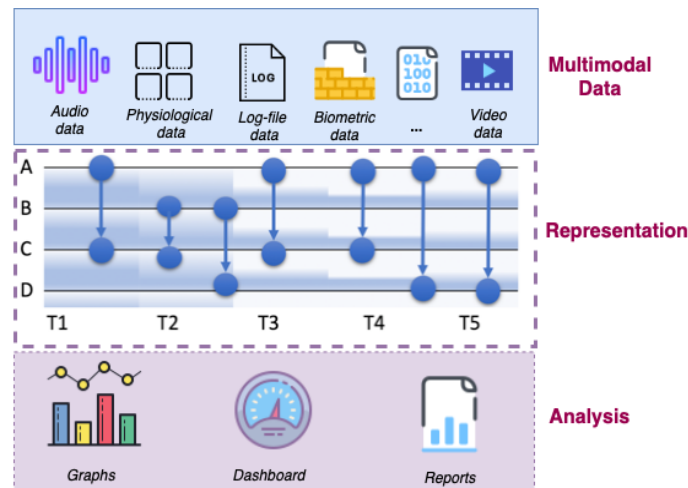
MMLA uses multiple synchronized sensing modalities to record learners' interactions, spatial data, physiological indicators as well as eye- and body movements. For example, physiological measurements such as heart rate data can be linked to certain learner's experiences (Ochoa & Worsley, 2016). Multimodal data can be recorded in real-time and amass unprecedented volumes of high resolution learner temporal data. As researchers try to make sense of these complex data, they have used several approaches for analysis either separately or in combination. Such approaches include traditional statistics, machine learning and qualitative methods (Viberg, Hatakka, Bälter, & Mavroudi, 2018). The complex interactions among learners - and learning resources - were earlier studied using well established network representations (Cela, Sicilia, & Sánchez, 2014), which employ network methods (i.e., powerful tools for the study of the relational data). They have been used successfully by educational researchers to for example, intuitively map interactions in simple understandable visual graphs, to reveal the structural dynamics of groups of learners, and to identify roles and influencers in a collaborative environment (Cela et al., 2014). To represent the relations as a network, researchers often aggregate all interactions in what is known as an 'aggregate' or static network (i.e., a compilation of all interactions). In doing so, the static network representation ignores the time aspect, considers that relations are permanent, and disregards the dynamics of the represented interaction process and related learning activities (Holme, 2015). As such, static network representations are much limited in terms of a holistic understanding of learning as a continuous process occurring when students interact with: each other, teachers, the available learning resources and involved learning environments. Compressing the time dimension is reductionist and arguably simplistic. Earlier learning analytics studies have shown the importance of taking time into account when analyzing learning events (e.g., Chen, Resendes, Chai, & Hong, 2017; Malmberg et al., 2017; Molenaar & Järvelä, 2014; Saqr et al., 2019).

2 METHOD PROPOSAL

We argue that extending the current approach by retaining the temporal dimension and its related information is beneficial to: i) understand the continuous nature of the learning process, and ii) further suggest related actions aimed at improving student learning outcomes and relevant learner support and teaching. A multimodal temporal network analytical approach is thus believed to have the potential to help researchers to unravel the timeline of learning events, the sequence of interactions and the relational properties of the learning process; most importantly, its evolving nature. The captured multimodal data from multiple streams are both temporal and relational as they capture time-stamped interactions. Consequently, temporal networks could offer a solid model for representing multimodal data in meaningful ways. Nowadays, research in temporal networks methods have given rise to a growing set of visual and mathematical methods. Such methods have contributed to the understanding of complex phenomena such as information spread, modelling disease contagion and brain connectivity, for a review please see (e.g., Holme, 2015; Holme & Saramäki, 2012). For education, temporal network analysis of multimodal data offers powerful representations and modeling of the temporal dimensions (e.g., timing of interactions among learners and teachers, timing of interactions with learning resources, timing of interactions with learning environment/s) that underpin learning- and teaching processes. While other methods of temporal analysis, such as using time series analysis offer a rich tool set for temporal analysis, they

do not fully cover the relational continuous nature of interactions in a learning environment. Nonetheless, both methods are complimentary, and recent research is exploring methods to combine the strengths of each method. We propose a three step approach to Multimodal Temporal Network Analysis to improve learner support and teaching. Such an approach is suggested to include three key mutually constituting parts: data, representations and analysis (Figure 1)

Figure 1: Multimodal Temporal Network Analysis



Data

- Multimodal data: spatial and proximity data
- Audio data and discourse capturing
- Video data
- Log-file data
- Physiological measurements such as eye movement, electro-dermal activity (galvanize skin response)

Representation

Networks enable the representation and modeling of the collected data

- proximity, audio, computer mediated interactions and spatial data be represented as networks of interactions among learners
- proximity, eye interaction with the elements of learning environment such as equipment, artefacts or laboratory tools could be represented as affiliation networks.
- physiological data:
 - as networks of physiological synchronization among collaborators
 - physiological data such as heart rate could be incorporated as edge weights or signs.

Analysis

- Temporal networks methods offers several models for the visualization (i.e., learner-and teacher support mechanisms) and the mathematical analysis of networks such as the spread of information, the evolution of communities, influencers, and the key drivers of the process.

Future research directions

By applying multimodal temporal network analysis, we suggest that we can better understand multifaceted aspects of temporal learning processes occurring in learners' interactions with each other and/or teachers, as well as the interactions with the involved learning environments and learning resources in use.

Some examples of potential research questions that could be addressed include:

- How can we understand the social regulation of collaborative learning that unfolds and develops over time?

- How do successful teams of learners manage learning tasks, and what characterizes a successful team process?
- Can temporal network representation offer an accurate model for the understanding of group dynamics, and if so, how?
- What are the temporal characteristics of productive collaboration considering the interplay between stress levels (biometric data), communication (audio or text), and eye-movements (video)?

Example: capturing multimodal data of a group of learners, audio data can be used to obtain a network of students' interactions; eye tracking and video data could be used to obtain another network of eye contact; physiological sensors could be used to capture levels of physiological arousal. Mapping these multiple signals together one could understand the interactions that lead to successful social regulation of teamwork, when they happened and how they progressed.

All in all, we propose to develop and adopt a new methodological approach for MMLA research, the Multimodal Temporal Network Analysis that, on the one hand, incorporates temporal aspects of learning as an analytical lens in order to capture learning as a continuous process, and on the other hand, combines it with network analysis as an analytical method in order to also capture the interdependent nature of learning interactions. By doing so, we argue that MMLA research is enhanced with a stronger ability to represent and model the complex interdependent multimodal learning interactions and processes that take place in space as well as in time.

REFERENCES

- Cela, K. L., Sicilia, M. Á., & Sánchez, S. (2014). Social Network Analysis in E-Learning Environments: A Preliminary Systematic Review. *Educational Psychology Review*, 27(1), 219–246. <https://doi.org/10.1007/s10648-014-9276-0>
- Chen, B., Resendes, M., Chai, C. S., & Hong, H.-Y. (2017). Two tales of time: uncovering the significance of sequential patterns among contribution types in knowledge-building discourse. *Interactive Learning Environments*, 25(2), 162–175. <https://doi.org/10.1080/10494820.2016.1276081>
- Holme, P. (2015). Modern temporal network theory: a colloquium. *European Physical Journal B*, 88(9). <https://doi.org/10.1140/epjb/e2015-60657-4>
- Holme, P., & Saramäki, J. (2012). Temporal networks. *Physics Reports*, 519(3), 97–125. <https://doi.org/10.1016/j.physrep.2012.03.001>
- Järvelä, S., Kirschner, P. A., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., ... Järvenoja, H. (2014). Enhancing socially shared regulation in collaborative learning groups: designing for CSCL regulation tools. *Educational Technology Research and Development*, 63(1), 125–142. <https://doi.org/10.1007/s11423-014-9358-1>
- Malmberg, J., Järvelä, S., & Järvenoja, H. (2017). Capturing temporal and sequential patterns of self-, co-, and socially shared regulation in the context of collaborative learning. *Contemporary Educational Psychology*, 49, 160–174. <https://doi.org/10.1016/j.cedpsych.2017.01.009>
- Molenaar, I., & Järvelä, S. (2014). Sequential and temporal characteristics of self and socially regulated learning. *Metacognition and Learning*, 9(2), 75–85. <https://doi.org/10.1007/s11409-014-9114-2>
- Ochoa, X., & Worsley, M. (2016). Augmenting Learning Analytics with Multimodal Sensory Data. *Journal of Learning Analytics*, 3(2), 213–219. <https://doi.org/10.18608/jla.2016.32.10>
- Saqr, M., Fors, U., & Nouri, J. (2019). Time to focus on the temporal dimension of learning: a learning analytics study of the temporal patterns of students' interactions and self-regulation. *International Journal of Technology Enhanced Learning*, 11(4), 398. <https://doi.org/10.1504/ijtel.2019.10020597>
- Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior*, 89(July), 98–110. <https://doi.org/10.1016/j.chb.2018.07.027>

Using Multimodal Learning Analytics to Explore how Children Experience Educational Motion-Based Touchless Games

Serena Lee-Cultura, Kshitij Sharma, Michail Giannakos

Norwegian University of Science and Technology (NTNU), Trondheim, Norway
serena.leecultura@ntnu.no

ABSTRACT: Leveraging motion-based touchless games (MBTG) to support children's learning is appealing and technically challenging. The application of multimodal learning analytics (MMLA) can help researchers to better understand how children experience learning through movement by providing insights into children's cognitive, behavioural, interaction, and learning processes. However, there is limited knowledge about exploiting the integration of MMLA into the use of educational MBTG in children's learning. We present an in-progress study in which we conducted an experiment with 55 children, playing three different educational MBTG centred on the development of math and English competencies. We collected multimodal data from 6 different sources: eye-tracking glasses, video, wristband, game analytics, Kinect point cloud, and questionnaires. Future analysis will explore relationships between the various multimodal data, in pursuit of establishing a more holistic understanding of children's cognitive, behavioural, interaction, and learning processes experienced while engaged with MBTG for learning.

Keywords: Motion-Based Games; Multimodal Learning Analytics; Educational Technologies; Child-Computer Interaction; Embodied Learning.

1 INTRODUCTION AND MOTIVATION

Researchers are looking for new ways to engage children in the learning process (Yap, Zheng, Tay, Yen, & Do, 2015), and better understand child-computer interactions in educational contexts. On the one hand, recent studies demonstrate support for Motion-Based Touchless Games (MBTG) as a potential pedagogical instrument capable of transforming the learning experience through amplified student motivation and engagement (Hsu, 2011; Kourakli et al., 2017). On the other hand, the application of multimodal learning analytics (MMLA) equips researchers with a more holistic understanding of the learning process (Blikstein & Worsley, 2016), specifically regarding learner-computer interactions (Giannakos, Sharma, Pappas, Kostakos, & Velloso, 2019). However, despite the wealth of potential advantages that may arise from the integration of multimodal data (MMD) capture when utilising educational MBTG, there is limited work exploiting the combination of these powerful tools.

We present an in-progress study that captures MMD during children's interactions with educational MBTG centred on the development of maths and English competencies. We discuss the study's overarching objective, research design, work currently completed, and future directions for analysis. The aim of this submission is to provide example research to serve as the centre for discussion on ways MMLA might be used to advance understanding of children's learning via MBTG. We hope to exchange ideas for potential analysis not currently under consideration, identify possible collaborations with interested parties, as well as encourage others to adopt this exciting area of research.

2 RESEARCH QUESTIONS

The central objective of this research is to explore how children experience educational MBTG. We aim to understand the cognitive, behavioural, and interaction processes experienced in this context and to investigate how these processes relate to the children's learning, acceptance and perception of MBTG games, their interaction modes and knowledge to be acquired. Specifically, we suggest that when answering questions as part of the MBTG learning experience, children undergo the See-Solve-Move-Select cycle (SSMS). During the SSMS cycle, children (1) see and understands the problem, (2) solve the problem mentally, (3) move their body to initiate the selection process, and finally (4) select and manipulates their answer via gestural interaction (see Figure 1). Using MMD capture, we aim to investigate the processes that occur during the different phases of the SSMS cycle, in pursuit of obtaining a holistic representation of the children's learning experience.

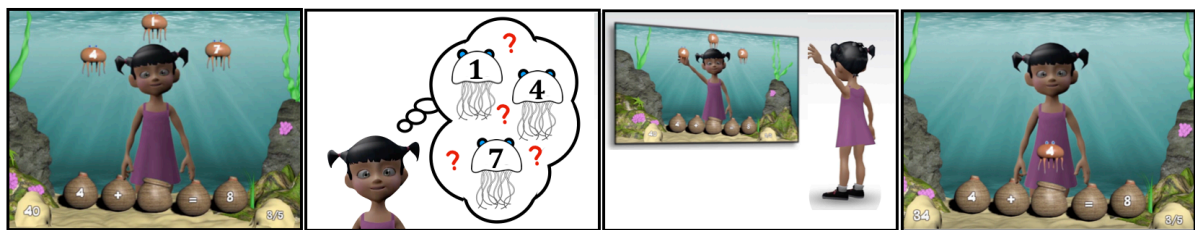


Figure 1. The 4 stages of the See-Solve-Move-Select cycle. Left: the child initially sees and must understand the problems. Middle left: The child solves the problem mentally. Middle right: The moves to initiate the selection process. Right: The child selects their answer.

3 RELATED WORK

Though technological advancements have only recently enabled the emergence of Motion-Based Touchless (MBT) devices, their application in education has seen much traction, with research permeating maths (Johnson, Pavleas, & Chang, 2013; Smith, King, & Hoyte, 2014) and language development (Yap et al., 2015). Notable studies suggest that in the context of maths, MBTG might have a positive impact on student learning; particularly concerning enhanced problem understanding (Smith et al., 2014) and increased academic performance (Kourakli et al., 2017; Tsai, Kuo, Chu, & Yen, 2015). MBT technology has also shown promise in development of language skills. For example, the Word Out! system (Yap et al., 2015) used motion sensing to aid children in learning to recognise the characteristic features of the alphabet. Results showed that the system motivated children, while fostering creative and collaborative strategies throughout their playful educational experiences. Collectively, these contributions demonstrate that researchers and teachers are beginning to consider MBT technology as a viable solution by which to augment the current instructional approach (Hsu, 2011). However, research shows that the criteria used to assess children's experience with MBT in the context of learning maths and English is mainly centred on subjective measures, such as motivation (Tsai et al., 2015; Yap et al., 2015) and enjoyment (Tsai et al., 2015). In short, researchers are not exploiting the full capacities of MMD to assess student's learning experiences. However, recent studies suggest that MMLA are capable of providing deep evaluations of students experiences across different learning contexts (Blikstein & Worsley, 2016; Worsley & Blikstein, 2015). Researchers have exploited MMLA to identify predictors for student performance and behaviour in adaptive learning environments (Sharma, Papamitsiou, & Giannakos, 2019), and better understand the collaborative

process of pair programming tasks in children's education (Grover et al., 2016). That being said, MMD capture has not been widely adopted in the field of learning analytics (Blikstein & Worsley, 2016). Consequently, there is limited knowledge exploiting the use of MMLA to better understand and assess the processes which occur during the use of MBT technologies in children's education. Accordingly, we identify the need for research to adhere to a fuller arsenal of data collection and assessment tactics (i.e., MMLA) in pursuit of developing a deeper understanding of the synergy between children's engagement with educational MBTG and the processes associated with their learning outcomes.

4 GAMES

Our study used three adaptable educational MBTG games from a commercial Kinect-based platform: Suffiz, Marvy Learns, and Sea Formuli. Each game was single player and focused on the development of math or English skills. Children interacted with the game content by moving their bodies and performing single hand mid-air hand gestures to move items on-screen. Though the focus of each game differed (Suffiz concentrated on English skills, Sea Formuli centred on arithmetic, and Marvy Learns targeted geometry or English depending on the grade setting), all of the questions presented were structured as either a multiple-choice question or a sorting problem. Furthermore, the way that each child interacted with the game content was identical across the game play sessions. That is, to answer a question, the child needed to use a pre-defined gestural selection mode (i.e., a delay or a grab motion) to select the correct item from a collection of items and then move the selected item to a target destination. Both the delay and grab gestures were single hand movements and only recognised when performed by the child's dominant hand. The delay gesture required the child to raise their hand, with palm facing forward, and hold it stable for a 1.5s. As the delay selection was progressing, visual feedback was displayed to the user. The grab gesture required the player to produce and maintain a grabbing gesture. In both cases, once the item was selected, it followed the child's hand movement. Moreover, each game took a different approach to player representation within the game (i.e., level of immersion). In Suffiz, a hand shaped cursor tracked the movement of the player's hand (low level immersion). In Marvy Learns, the player's full body movement was mapped to a creature avatar (medium level immersion). Finally, in Sea Formuli, a video image of the player projected the child's full body into the game setting (high level of immersion), see Figure 2.



Figure 2: Player representations of the three games. Left: Suffiz represents the player via hand cursor. Middle: Marvy Learns maps the player's full body to an avatar. Right: Sea Formuli inserts a video of the player inside the game.

5 METHODS

5.1 Context

The context of our experiment takes place in two different venues. Namely, a children's science centre and an elementary school in a European country. In both cases, researchers were present onsite during the game play sessions to assist children in understanding game play and gesture execution.

5.2 Participants

Our sample was composed of 55 elementary school children with an average age of 10.27 years (min = 8, max = 11 years). 25 of the children were female and 30 were male. All of the children were typically developing. Furthermore, each child participated in 9 games play sessions (3 consecutive rounds of each of the aforementioned games) in the science centre or elementary school setting.

5.3 Procedure/Experiment

We conducted a four-phase within-between groups experiment to investigate the learning, behavioural and interaction processes experienced by children as they engaged with educational MBTG centred on developing math and English competencies (see Figure 3). The level of immersion (i.e., cursor, full body avatar mapping, and video of self) was the within-groups condition and the selection mode (i.e., delay, grab) was the between groups condition. We balanced the assignment of selection modes and the order of level of immersion (i.e., order in which the games were played). After obtaining parental written consent, children were given a pair of Tobii eye-tracking glasses, and an Empatica E4 wristband to wear (phase 1). Then, for each game, children played three consecutive sessions: a practice round, in which researchers assisted the child in understanding the game's objective and rules (phase 2), and two non-practice sessions (phase 3). Finally, children filled out a questionnaire. None of the children had prior experience with MBT technologies, or Kinect games.

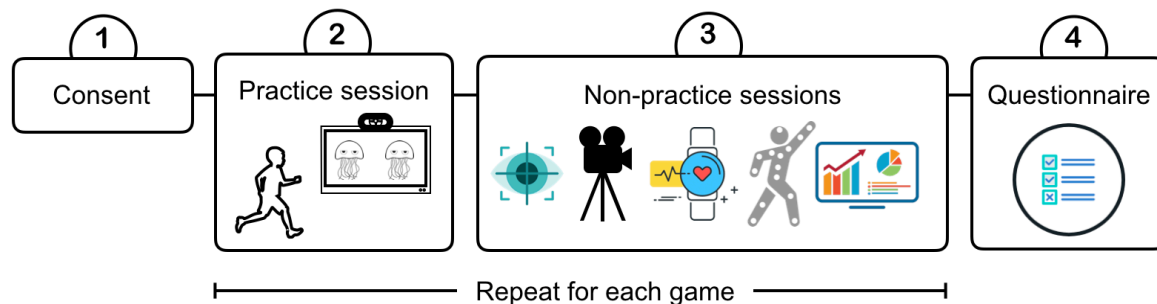


Figure 3 The four stages of the experimental study.

5.4 Multi Modal Data Collection

We collected six different data sources from each child throughout the duration of the sessions.

Eye tracking: Children wore Tobii eye-tracking glasses allowing us to capture both the eye-tracking data and the children's field of view (objective camera on the nose-bridge).

Facial Video: We captured children's facial expressions using a LogiTech HD web camera situated on top of the screen and directed at the child. The web camera was set to 200% zoom to enable clear capture of the child's face.

Wrist Band: Participants wore an Empatica E4 wristband, from which we recorded 4 different measurements: 1) HR at 1Hz, 2) EDA at 64 Hz, 3) body temperature at 4 Hz, and BVP at 4 Hz.

Kinect Skeleton: We collected the complete skeletal data provided by Kinect Point Cloud. Specifically, this includes information on the child's joint movement (i.e., joint orientation, depth position), collected at successive time fixed intervals.

Game Analytics: We collected system log files containing event time stamps corresponding various child-computer interactions, such as when an item is selected and released or when a question is answered. As well, a report outlining various performance metrics, such as the child's correctness and reaction time, was also obtained per session.

Questionnaire: This included basic demographic data, such as the child's age, gender, and school grade, as well as 14 5-point Likert scale questions addressing their experience and emotions.

6 CONCLUSION AND FUTURE DIRECTIONS

Our aim is to better understand how children experience educational MBTG for maths and English, by identifying and examining the cognitive, behavioural, and interaction processes that occur in learning. Specifically, we plan to investigate our proposed SSMS cycle, and how it relates to children's learning, acceptance and perception of MBTG games, their selection modes, and the knowledge acquired. Our ongoing work on this experiment will exploit the use of MMLA. Narrowing the research scope considerably, we start by asking how the level of immersion in educational MBTG relates to student affect and behavioural processes. Our upcoming analysis will employ data captured from eye-tracking glasses, wristbands, video and Kinect sensor, to examine the relationships between various aspects of children's embodied learning experience, such as levels of stress, arousal, fatigue, cognitive load, global and local information processing, on-task/off-task ratio, facial expression and amount of bodily movement. We hope such analysis may scaffold the understanding of processes that occur during children's interactions with MBTG in educational contexts. As MBT technologies continue to establish themselves as rich resources for creating meaningful interactions in children's education, we highlight the importance of this work's relevance to the LAK community, in terms of exploring the design and assessment of learning experiences via MBTG.

REFERENCES

- Blikstein, P., & Worsley, M. (2016). Multimodal Learning Analytics and Education Data Mining: using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, 3(2), 220-238.
- Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., & Velloso, E. (2019). Multimodal data as a means to understand the learning experience. *International Journal of Information Management*, 48, 108-119.

- Grover, S., Bienkowski, M., Tamrakar, A., Siddiquie, B., Salter, D., & Divakaran, A. (2016). *Multimodal analytics to study collaborative problem solving in pair programming*. Paper presented at the Proceedings of the Sixth International Conference on Learning Analytics & Knowledge.
- Hsu, H.-m. J. (2011). The potential of kinect in education. *International Journal of Information and Education Technology*, 1(5), 365.
- Johnson, K., Pavleas, J., & Chang, J. (2013). Kinecting to mathematics through embodied interactions. *Computer*, 46(10), 101-104.
- Kourakli, M., Altanis, I., Retalis, S., Boloudakis, M., Zbainos, D., & Antonopoulou, K. (2017). Towards the improvement of the cognitive, motoric and academic skills of students with special educational needs using Kinect learning games. *International Journal of Child-Computer Interaction*, 11, 28-39.
- Sharma, K., Papamitsiou, Z., & Giannakos, M. (2019). Building pipelines for educational data using AI and multimodal analytics: A “grey-box” approach. *British Journal of Educational Technology*.
- Smith, C. P., King, B., & Hoyte, J. (2014). Learning angles through movement: Critical actions for developing understanding in an embodied activity. *The Journal of Mathematical Behavior*, 36, 95-108.
- Tsai, C.-H., Kuo, Y.-H., Chu, K.-C., & Yen, J.-C. (2015). Development and evaluation of game-based learning system using the Microsoft Kinect sensor. *International Journal of Distributed Sensor Networks*, 11(7), 498560.
- Worsley, M., & Blikstein, P. (2015). *Leveraging multimodal learning analytics to differentiate student learning strategies*. Paper presented at the Proceedings of the Fifth International Conference on Learning Analytics And Knowledge.
- Yap, K., Zheng, C., Tay, A., Yen, C.-C., & Do, E. Y.-L. (2015). *Word out!: learning the alphabet through full body interactions*. Paper presented at the Proceedings of the 6th Augmented Human International Conference.

The challenge of interaction assignment on large digital tabletop displays for learning analytics

Matthias Ehlenz, Vlatko Lukarov, Ulrik Schroeder

Learning Technologies Research Group, RWTH Aachen University
[ehlenz,lukarov,schroeder]@informatik.rwth-aachen.de

ABSTRACT: One system, four learners, eight hands. A typical situation in collaborative learning with digital tabletops and a beautiful testbed for learning analytics if not for the fact, that a central question remains open. Who did what? The first part of any xAPI-statement is the actor, which from the systems perspective can be any one of the current learners. This contribution describes the practical experience to employ learning analytics strategies in this non-traditional scenario. Collaborative learning on digital tabletops always involves multiple users interacting simultaneously with the system. Identifying with user is responsible for which interaction is challenging. Various approaches have been taken and are briefly described in this contribution.

Keywords: interaction assignment, multi-touch, digital tabletops, collaborative learning, machine learning

1 LEARNING ANALYTICS IN FACE-TO-FACE COLLABORATION

Science has come a long way in a lot of research areas and modern technologies take advantage of those achievements by combining the benefits and insights to create new or improve known solutions. Some of these areas are learner modelling [1], intelligent tutor systems [2] and learning analytics. Results surmount i.e. in adaptive feedback technologies, giving each user the most valuable feedback for her personality as well as her current situation.

Traditionally learning analytics aims at analysis of user behavior in learning processes to understand and improve those processes. A lot of steps have to be taken, a lot of questions to be asked. A good way to identify those questions is to adhere to the Learning Analytics Reference Model [3] and ask what data to gather, why do it at all, how to analyze and who is stakeholder and thus interested in the outcome. A question usually not asked is “Who is the learner?”, this time not from a psychological perspective but from a very practical point of view.

Multitouch tabletop displays got bigger and more affordable in the recent years and consequently found their ways into the educational system. Large display real estate allows multiple learners to interact with a single system at the same time thus bringing face to face collaboration back into technology-enhanced group learning processes. Multiple learning games and applications have been designed for research purposes at our institution, the most prominent a game to rehearse and practice regular expressions by dragging matching words into players zones of regular expressions.

The details of the game, its didactical approach and first findings can be found in [4], this contribution focuses on technical challenges and approaches of interaction assignment in that game.

1.1 The Problem of Assignment

Initially the problem of interaction assignment wasn't expected at all. The learning game, in our experimental setup played on an 84" tabletop display, provided each player with a regular expression of identical structure but different characters in each player's personal area directly in front of her. The common area in the middle featured "word cards" with words matching one of those expressions or not. The idea has been that each player drags matching cards into her area and the group is awarded points for each matching word, to achieve higher scores everyone needs to get involved, the collaboration intended to be verbally, learners explaining each other the structure of current levels regular expression.

In fact, this behavior is present and observed in nearly every session. What has been unaccounted for is behavior of a different kind. In the first implementation the learning analytics component attributed the drag and drop interaction to the player standing at the position of the regular expressions drop area. Practice showed two behavioral patterns not expected and therefore not covered by this naïve assumption. Pattern A is referenced in the following as "Sorting behavior". Some learners tend to sort more than others, but it is observed in all subjects so far to a varying degree. Interaction does not necessarily end in a target zone. The motives differ, some pull word cards closer, either to read or claim temporary possession, some pre-sort in "might fit – definitely doesn't match", some pre-sort for the whole group. The last motive is pretty close to pattern B: Players pushing cards which they might think will fit in other players "drop zone" close to them, leaving the final decision to them. Not all players show that inhibition to "intrude" their colleagues' personal space, making up pattern B, which will be consecutively called "Cross interaction", learners taking a word card and dropping them into one of the other three target areas not in front of themselves.

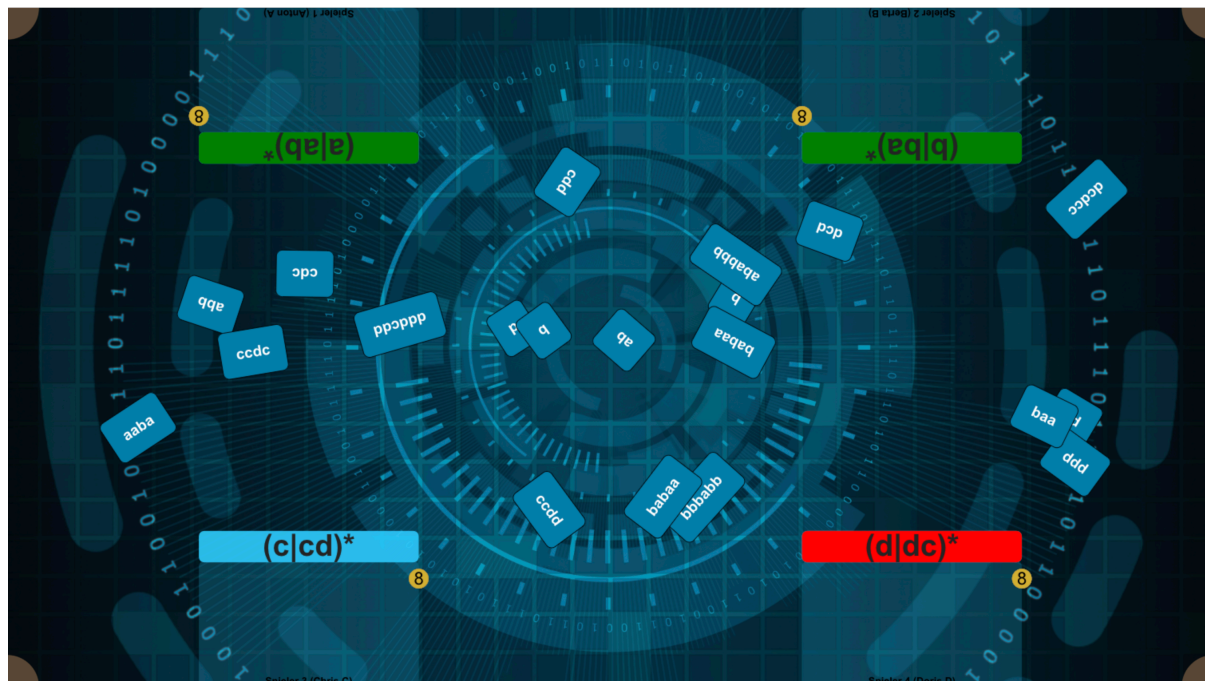


Figure 1: The Game "RegEx"

Both patterns are challenging from a learning analytics perspective. The touchscreen technology cannot differentiate between disparate fingers on its surface and thus not tell which player interacted with which element. Pattern A resulted in simply unassigned interactions, Pattern B even in falsely assigned interactions, both dangerous for a complete attempt on learning analytics.

In the following years, several attempts have been taken to tackle this challenge and will be described in the following:

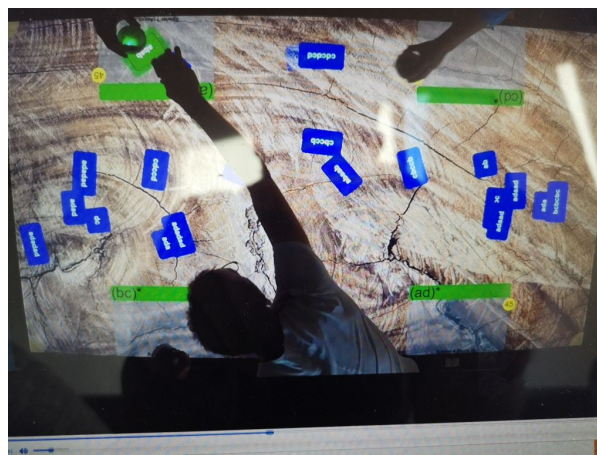
1.1.1 Active Tangibles

The first approach came, in a way, naturally, since the learning game was developed in the context of a publicly funded research project, TABULA, which focused on tangible learning with graspable elements. In this project, active tangibles have been developed, small handheld devices which could interact with the system by providing a recognizable pattern on the bottom side as well as establishing a Bluetooth connection to the application. Apart from offering new interaction mechanisms and feedback channels, such devices can be uniquely identified by the system and thereby identify the user.

The idea in general holds up. Tangibles can help to identify users and provide reliable data for learning analytics. Nonetheless there are arguments against further application for this approach: First, this technology is still prototypical. Just that we can use it does not solve the problem itself, it's like a crutch not available to everyone. Second, and this is more serious, the project showed that tangibles change user behavior and therefore does not lead to valid results on the effect of multitouch collaborative learning and cooperative behavior in general. I.e. some participants complained to have felt impaired as they have been told to only use the tangible, leaving one hand unused.

1.1.2 Motion Tracking

Often suggested, there was a preliminary test regarding the suitability of motion tracking devices like Microsofts Kinect, which was available for PC for a slight time frame. While the idea of skeleton tracking is appealing, it was not possible to find any angle suitable for this application. Front view is obstructed by the physically opposing players and could not be altered without changing the experimental setup and thereby the user's behavior in itself. Top view had significant problems both with skeletal tracking (due to highly unusual perspective) as well as obscured view of hands.



After this first trials the approach did not look promising at and it has been decided to drop it completely and invest resources elsewhere.

1.1.3 Manual Labeling

The last resort for every researcher. Manual labelling. A test run was conducted in which the test manager was instructed to take notes of observed behavior of patters A or B. This proved to be unpractical due to a pretty high frequency of such events and the inability of a single person to follow the actions of four people simultaneously. Consequently, the experimental setup has been changed and wide-angle cameras have been installed above the tabletop display.

The video stream proved to be helpful, but manual labelling has been time-consuming and sometimes difficult as finding a single interaction in a set of usually around 300 events in a five-minute-session and the corresponding video material is difficult and sometimes, in case of multiple simultaneous interactions by different players, close to objectively impossible.

1.1.4 Tool-supported post-processing

The difficulties in the manual labeling process led to the development of a dedicated tool for interaction assignment which streamlined the process by several orders of magnitude. The video file is loaded into the tool, the dataset of the session according to the file name is fetched from the server and video and event stream are synchronized by the click of a single button. After this, the application workflow is as follows: First all “invalid” data is filtered (by our definition events with a duration below 200ms as those are mostly artifacts and noise), then all unassigned events are showed and manually assigned to one of the four players. Finally, all events for each player is played back and either confirmed or corrected by the post processing user.

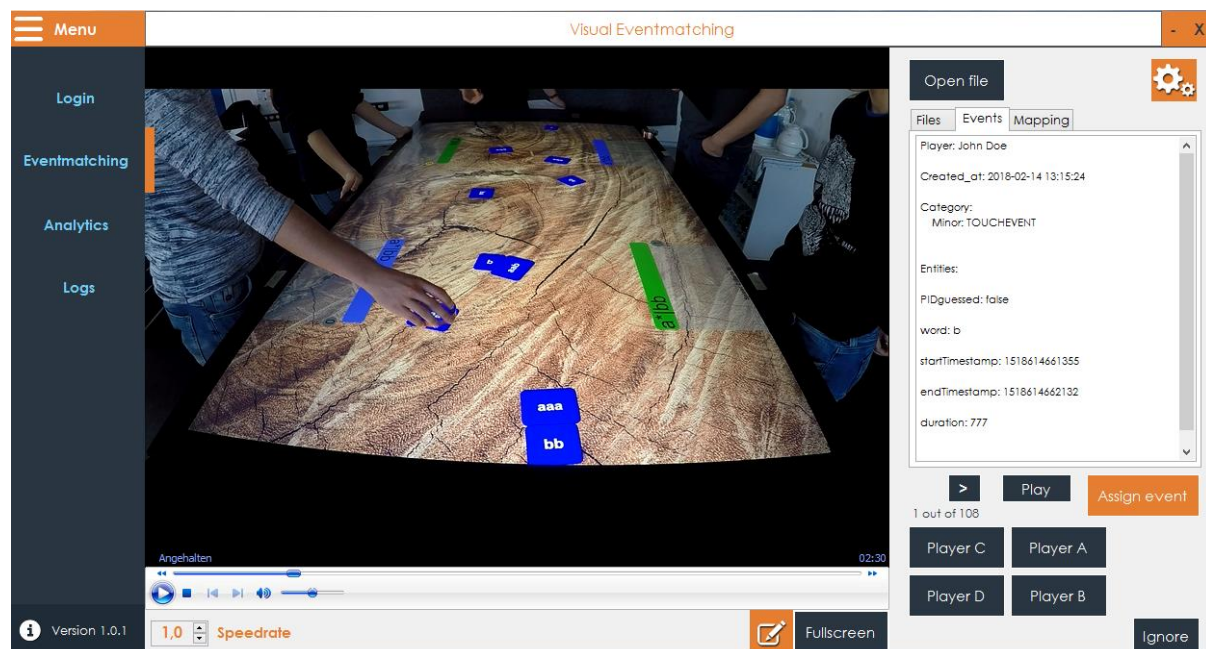


Figure 2: Tool-supported labelling

Events are showed in a bigger frame as video stream starting at the events timestamp and can be slowed down and replayed by press of a button. A small frame on the upper right contains a small rendering of the interaction's movement and below the events meta-information is provided.

Finally, the results can be saved back to the server and exported as .csv file. Pre-labelling is done solely on drop-zone information. This tool improved the post processing significantly, labelling of a five-minute session of about 300 events dropped from several hours to below 30 minutes for an experienced researcher.

1.1.5 Machine-learning approach

Various ideas have been developed to further improve the pre-labelling or even automated assignment of interactions to learners. The most promising has been the evaluation of current machine learning algorithms to the data.

Fundamental to this is the idea that there are several features and characteristics depending on the interacting persons position when standing at such a large tabletop that prediction of that position from speed, acceleration, angle and curvature of the interaction seemed feasible.

The idea and feature calculation showed similarities to handwriting recognition, so a bachelor thesis¹ evaluated the usage of the recommended Support Vector Machines, Random Forests, AdaBoost, CNNs and RNNs on a labeled training set of about 4000 interactions.

Table 1: Comparison of Algorithms

Algorithm	Avg. Accuracy
SVN	70.44
RandomForest RandomForestClassifier	- 81.7
RandomForest RandomForestClassifier	- 83.3
AdaBoost	77.16
CNN	78.7
RNN	80.54

The training data set proved to be far too small for most of those algorithms, but first results suggest that Random Forest and Recurring Neural Networks look most promising with accuracy above 80%. While this is still far from automated assignment with a sufficient degree of certainty it brings further improvement to the post-processing of the interaction data by pre-labeling the data before the manual check. The algorithms will be reevaluated as the labeled dataset grows.

¹ <http://publications.rwth-aachen.de/record/764115>

2 FINDINGS AND CONCLUSION

Correct attribution remains our biggest challenge in collaborative learning analytics. We strive to gather more data and intensify our machine learning efforts but look in different directions as well, starting with usage of eye-tracking glasses this semester.

REFERENCES

- [1] Bull, S., Kickmeier-Rust, M., Vatrpu, R. K., Johnson, M. D., Hammermueller, K., Byrne, W., ... & Meissl-Egghart, G. (2013, September). Learning, learning analytics, activity visualisation and open learner model: Confusing?. In *European Conference on Technology Enhanced Learning* (pp. 532-535). Springer, Berlin, Heidelberg.
- [2] Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of educational research*, 86(1), 42-78.
- [3] Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2013). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5-6), 318-331.
- [4] Ehlenz, M., Leonhardt, T., Cherek, C., Wilkowska, W., & Schroeder, U. (2018, November). The lone wolf dies, the pack survives? Analyzing a Computer Science Learning Application on a Multitouch-Tabletop. In *Proceedings of the 18th Koli Calling International Conference on Computing Education Research* (pp. 1-8).

Facilitating Self-Regulated Learning with Personalized Scaffolds on Student's own Regulation Activities

Joep van der Graaf¹, Inge Molenaar¹, Lyn Lim², Yizhou Fan³, Katharina Engelmann²,
Dragan Gašević^{3,4}, Maria Bannert²

¹Behavioural Science Institute, Radboud University, The Netherlands

²Technical University of Munich, Germany

³University of Edinburgh, Edinburgh, UK

⁴Monash University, Melbourne, Australia

j.vandergaaf@pwo.ru.nl

ABSTRACT: The focus of education is increasingly set on students' ability to regulate their own learning within technology-enhanced learning environments. Scaffolds have been used to foster self-regulated learning, but scaffolds often are standardized and do not adapt to the individual learning process. Learning analytics and machine learning offer an approach to better understand SRL-processes during learning. Yet, current approaches lack validity or require extensive analysis after the learning process. The FLORA project aims to investigate how to advance support given to students by i) improving unobtrusive data collection and machine learning techniques to gain better measurement and understanding of SRL-processes and ii) using these new insights to facilitate student's SRL by providing personalized scaffolds. We will reach this goal by investigating and improving trace data in exploratory studies (exploratory study1 and study 2) and using the insight gained from these studies to develop and test personalized scaffolds based on individual learning processes in laboratory (experimental study 3 and study 4) and a subsequent field study (field study 5). At the moment study 2 is ongoing. The setup consists of a learning environment presented on a computer with a screen-based eye-tracker. Other data sources are log files and audio of students' think aloud. The analysis will focus on detecting sequences that are indicative of micro-level self-regulated learning processes and aligning them between the different data sources.

Keywords: self-regulated-learning; instructional scaffolds; personalized learning; learning analytics; machine learning; adaptive systems.

1 THE FLORA PROJECT

The FLORA project aims to improve measurement of self-regulated learning by using multimodal learning analytics. Self-regulated learning (SRL) occurs when learners monitor and regulate content they access and operations they apply to content as they pursue goals to augment and edit prior knowledge [1]. SRL is related to better learning outcomes and SRL interventions improve SRL and learning outcomes. Recently, the need for improved measurement of SRL has increased, because effects of interventions on actual SRL behavior were limited [2]. A solution is to assess SRL at a more fine-grained level by measuring micro-level SRL processes.

SRL consists of cognitive activities related to learning the content and meta-cognitive activities related to regulation. Sub-categories of cognition refer to student's strategic information processing during learning such as reading the information, repeating it as well as deeper information

processing like elaboration, and organization of information processed. The metacognitive activities include five categories: planning, goal specification, orientation, monitoring, reflection, and evaluation which refer to the postulated metacognitive activities during SRL. When zooming in on these categories, many micro-level processes might be detected, such as content evaluation and monitoring progress towards learning goals in the category: monitoring [3]. It is the aim to detect these micro-level processes in study 1 and improve detection of SRL by adding instrumentation tools to the learning environment in study 2.

2 STUDY 1: MEASURING SRL PROCESSES

The aim of study 1 was to measure micro-level SRL processes. A learning environment was presented to students. The task for the students was to learn about three topics and to write an essay. Before and after this task, students' knowledge about the topics was assessed. Preliminary results show that there is a significant learning gain. The challenge is to link the learning gain to SRL processes. The objective is to analyze each data source (think aloud, log data, and eye-tracking) and extract behaviors that are indicative of micro-level SRL processes, see Fig. 1 for a schematic overview.

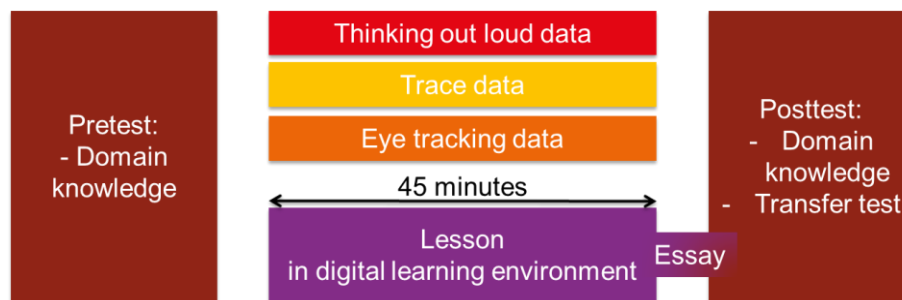


Fig. 1. A schematic overview of study 1.

3 STUDY 2: INSTRUMENTATION TOOLS

To improve traceability of SRL processes, instrumentation tools have been added in study 2. These include a timer, a note-taking and highlighting application, a planner tool, a search function, and a hybrid read-write mode of the essay in which both the text and essay is visible. These tools allow learners to reveal SRL processes, which should be reflected in improved traceability in think aloud, log, and eye-tracking data. Aside from the instrumentation tools the setup is the same as in Study 1.

4 SENSOR/DATA GATHERING SETUPS AND PROTOTYPES

Three types of data were gathered: log files (mouse and keyboard), audio, and gaze. To record data a mouse, keyboard, microphone, and eye-tracker were used. The participant was seated in front of a monitor with a screen-based eye-tracker, microphone, keyboard, and mouse. The stimuli were presented on this monitor. In future studies, feedback will be provided as well, see Fig. 2 for an overview of the technical infrastructure.

Audio was recorded to measure think aloud data. The participants were instructed to think aloud. Think aloud consisted of reading text, stating learning goals, mentioning the creation of notes,

stating navigational actions, etc.. To make sense out of the *log files*, logs have to be interpreted in the context of the learning environment. Mouse clicks mostly indicate navigation, while typing is most common for the note-taking function and the essay. *Gaze data* was recorded using a screen-based eye-tracker. The raw data consisted of a timestamp (sampling rate is 300 Hz), coordinates of the where the participants is looking, pupil dilation, Areas Of Interest (AOIs) if enabled, and more. For all data sources, a coding scheme is needed to label and analyze the data. Analysis will focus on sequences of actions that can be indicative of specific SRL processes.

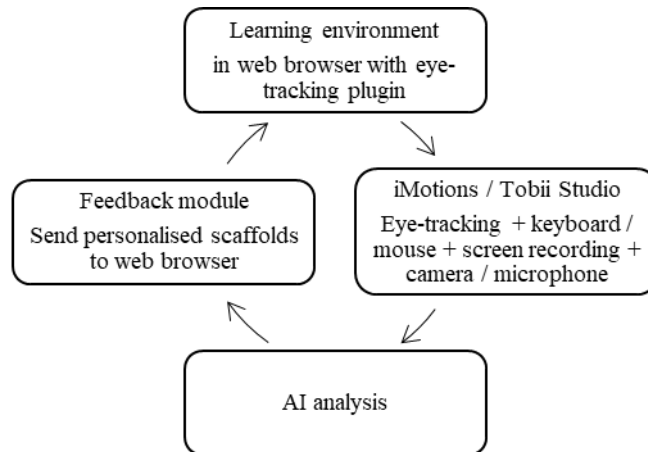


Fig. 2. An overview of the technical infrastructure.

5 THE WORKSHOP

The first part of the workshop consists of collecting data with the presented setup and a shorter task (15 minutes). In the second part, three groups will each investigate a single data source (think aloud, log data, or eye-tracking). The goal is to extract and analyze SRL processes and identify the value of instrumentation tools. To do so, data and a coding scheme will be provided. During this process, each group evaluates the data source in relation to detection of SRL. Advantages and disadvantages will be identified and discussed at the end when the groups come together to share the results.

REFERENCES

1. Winne, P.: Paradigmatic Dimensions of Instrumentation and Analytic Methods in Research on Self-Regulated Learning. *Computer in Human Behavior*, 96, 285-289 (2019)
2. Schunk, D.H., Greene, J.A.: *Handbook of Self-Regulation of Learning and Performance*. 2nd ed. Routledge, New York (2018).
3. Greene, J.A., & Azevedo, R.: A macro-level analysis of SRL processes and their relations to the acquisition of a sophisticated mental model of a complex system. *Contemporary Educational Psychology*, 34, 18-29 (2009).
4. Papamitsiou, Z., Economides, A.: *Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence*. *Educational Technology & Society*, 17(4), 49-64 (2014).

Endowing Head-Mounted Displays with Physiological Sensing for Augmenting Human Learning and Cognition

Evangelos Niforatos

Norwegian University of Science and Technology (NTNU), Trondheim, Norway
evangelos.niforatos@ntnu.no

Athanasios Vourvopoulos

Instituto Superior Técnico, University of Lisbon
athanasios.vourvopoulos@tecnico.ulisboa.pt

Michail Giannakos

Norwegian University of Science and Technology (NTNU), Trondheim, Norway
michailg@ntnu.no

ABSTRACT: The EEGlass prototype is a merger between a Head-Mounted Display (HMD) and a brain-sensing platform with a set of electroencephalography (EEG) electrodes at the contact points with the skull. EEGlass measures unobtrusively the activity of the human brain facilitating the interaction with HMDs for augmenting human cognition. Among others, EEGlass is intended for collection of context-aware EEG measurements, supporting learning and cognitive experiments outside the laboratory environment. Thus, we expect EEGlass will promote the implementation and application of ecologically valid research methods (studies in the user's natural context).

Keywords: Head-Mounted Displays, Electroencephalography, Brain-Computer Interfaces, Neuroadaptive Systems.

1 INTRODUCTION

Human cognition is typically a composite notion we use for describing the states of the cognitive processes that underpin it. Namely, attention; memory recall; learning; decision-making; and problem-solving. Augmenting human cognition boils down to gauging the states of the underlying cognitive processes and deciding on an intervention. On one hand, electroencephalography (EEG), and other measures of physiological responses, have been extensively utilized for measuring attention, monitoring cognitive workload, assessing learning experience, and even evaluating software usability. On the other hand, contemporary HMDs, such as Augmented Reality (AR) smart glasses, are progressively becoming socially acceptable and ubiquitous by approaching the size and design of normal eyewear (Niforatos & Vidal, 2019). Thus, a merger between HMDs and EEG appears to be promising. HMDs bear a significant potential in hosting an array of physiological sensors in contact with the human skull, while situated in front of our most highly-esteemed perceptive organ: our eyes. In this work, we draw on the HMD form factor for designing, developing, and evaluating EEGlass, an EEG-Eyewear prototype for ubiquitous brain-computer interaction (Vourvopoulos et al., 2019).

2 EEGGLASS PROTOTYPE

The latest version of the EEGlass prototype (see Figure 1) is comprised of a [Vuzix Blades](https://www.vuzix.com/products/blade-smart-glasses)¹ (Vuzix, Rochester, USA) HMD fitted with EEG electrodes that connect to a Cyton Biosensing Board by [OpenBCI](https://openbci.com/)² (OpenBCI, NY, USA). [Vuzix Blades](https://www.vuzix.com/products/blade-smart-glasses) is a pair of AR smart glasses that features a monocular and transparent waveguide display, with a 19-degrees field of view, and a resolution of 480 x 853 pixels. [Vuzix Blades](https://www.vuzix.com/products/blade-smart-glasses) is equipped with an 8MP camera, Bluetooth and Wi-Fi connectivity modules, a range of sensors (e.g., inertial measurement unit, microphones, etc.), and runs Android OS. [OpenBCI](https://openbci.com/) is a popular and relatively low-cost open hardware and software platform for the collection and analysis of biosignals such as EEG, EMG (Electromyography), and ECG (Electrocardiography), inspired by the grassroots movement of DIY (“Do It Yourself”). The Cyton board encompasses 8 biopotential input channels (for hosting up to 8 electrodes), a 3-axes accelerometer, local storage, Bluetooth connectivity module, while being fully programmable and Arduino compatible. Evidently, the EEGlass electrode topology is restricted by the form factor of [Vuzix Blades](https://www.vuzix.com/products/blade-smart-glasses) and at the contact points with the skull. Thus, EEGlass utilizes 3 electrodes (plus 2 for reference and ground) based on the standard 10-20 EEG system (see Figure 1) for measuring brain activity: 1 electrode placed inwards at the top of the eyewear bridge touching the skull at glabella, and 2 more electrodes at the inner side of the eyewear temples, touching the left and right mastoids, behind the left and right ears, respectively. Both the Cyton Board and [Vuzix Blades](https://www.vuzix.com/products/blade-smart-glasses) are connected to an external power source for enabling and prolonging mobile usage.

3 CURRENT STATE AND NEXT STEPS

Our first aim is to investigate how reliably EEGlass can capture brain activity, particularly when featuring an electrode topology imposed by the form factor of an HMD. For this, we compare brain activity captured via EEGlass with that captured via a standard EEG system as baseline. So far, we have tested a previous version of the EEGlass prototype, implemented with eyewear frames. Limited trials with 1 participant indicated that the EEGlass is capable of capturing brain activity manifested in two modes of resting state: (a) eyes open and focused on a target, and (b) eyes closed. Brain activity recorded during resting state with EEGlass demonstrated similar variations in frequency and amplitude to when recorded with an established EEG system. Recorded brain activity linked to upper limb motor-action displayed significant differences when compared to that captured with an established EEG system due to the fundamentally different electrode topology of EEGlass. Nevertheless, EEGlass managed to capture upper limb motor-action relying on signal propagation over the skull through volume conduction (van den Broek et al., 1998). EEGlass also detected subtle eye movements in 4 basic directions, displaying an eye-tracking potential particularly useful for navigating in HMD interfaces.

Low sample size (N=1) and stationary experimental settings are significant limitations that we will address over the next studies. However, human skull and brain anatomy is almost homogeneous,

¹ <https://www.vuzix.com/products/blade-smart-glasses>

² <https://openbci.com/>

and the HMD form factor ensures a rather stable electrode contact with the skull, only somewhat influenced by movement. In future iterations, we will utilize machine learning for training algorithms to match input from EEGlass to that of established EEG systems. We believe a merger between EEG and HMDs bears an unprecedented potential to “close the loop” by increasing the communication bandwidth between human and machine and paving the way for cognition-aware systems (Niforatos et al., 2017).

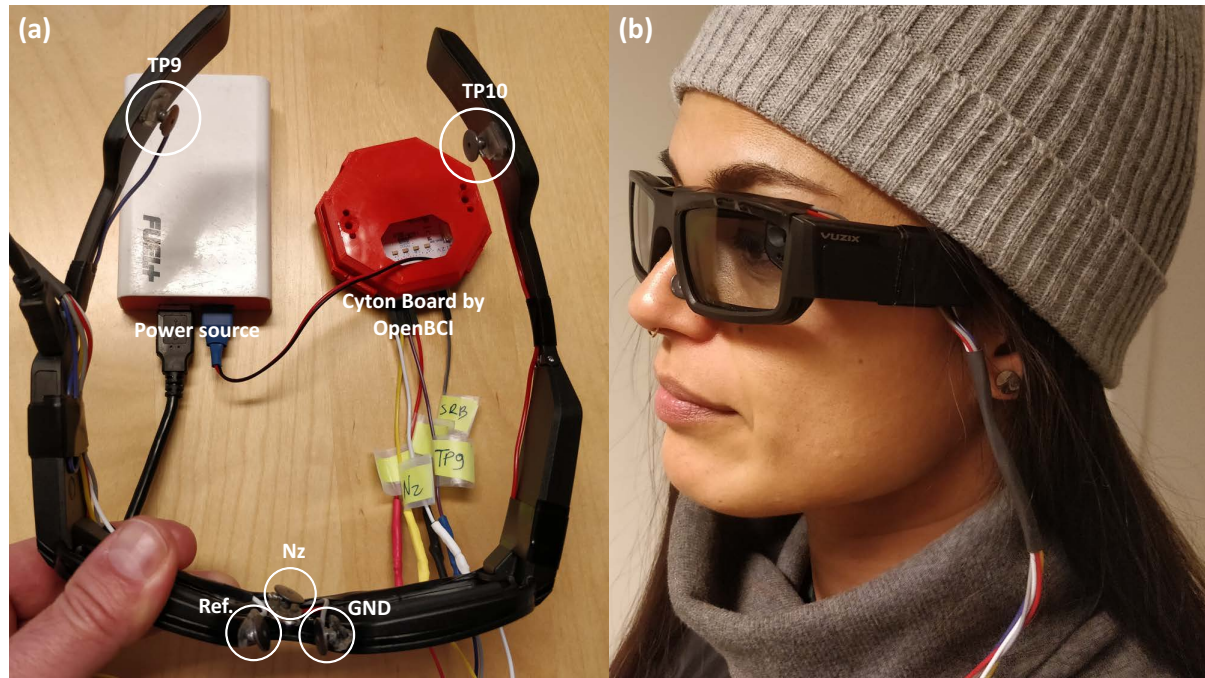


Figure 1: (a) The EEGlass prototype comprised of a Vuzix Blades HMD fitted with 5 EEG electrodes, based on the 10-20 system, connecting to a Cyton OpenBCI board and a mobile power supply. (b) A user wearing EEGlass.

4 APPLICATIONS FOR LEARNING

Besides the promising application areas in augmenting human cognition in general, we believe EEGlass also bears significant potential in facilitating learning. For example, after investigating EEGlass in reliably measuring cognitive activity in the wild, we will introduce it to the classroom. Although EEG can capture the subtle cognitive processes associated with learning (e.g., attention and concentration levels), performing EEG experiments in a classroom with the typical EEG headsets is deemed cumbersome and often inappropriate. Thus, we expect that EEGlass can be a viable alternative in collecting unobtrusively the brain activity of students related to learning. Moreover, the HMD component of EEGlass can be utilized for projecting information about the learning content in pre, post or during learning stage, and even on the go. We expect that by presenting our prototype to the CrossMMLA workshop, we will spark ideation and generate discussions about different applications and user scenarios for EEGlass about enhancing learning and the entire spectrum of human cognition.

ACKNOWLEDGEMENTS

This project is a product of a research collaboration between Norwegian University of Science and Technology (NTNU) and Institute for Systems and Robotics (ISR-Lisbon), receiving funding from Swiss National Science Foundation (SNSF) and the ERCIM “Alain Bensoussan” Fellowship.

REFERENCES

- van den Broek, S. P., Reinders, F., Donderwinkel, M., & Peters, M. J. (1998). Volume conduction effects in EEG and MEG. *Electroencephalography and clinical neurophysiology*, 106(6), 522-534.
- Niforatos, E., Vourvopoulos, A., & Langheinrich, M. (2017, September). Amplifying human cognition: bridging the cognitive gap between human and machine. In *Proceedings of the 2017 acm international joint conference on pervasive and ubiquitous computing and proceedings of the 2017 acm international symposium on wearable computers* (pp. 673-680). ACM
- Niforatos, E., & Vidal, M. (2019, March). Effects of a Monocular Laser-Based Head-Mounted Display on Human Night Vision. Retrieved from <https://dl.acm.org/citation.cfm?id=3311858>
- Vourvopoulos, A., Niforatos, E., & Giannakos, M. (2019, September). EEGlass: An EEG-eyeware prototype for ubiquitous brain-computer interaction. In *Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (pp. 647-652). ACM

Towards Collaboration Literacy Development through Multimodal Learning Analytics

Marcelo Worsley

Northwestern University, Evanston, IL, USA
marcelo.worsley@northwestern.edu

Xavier Ochoa

New York University, New York, NY, USA
xavier.ochoa@nyu.edu

ABSTRACT: The last ten years has involved significant growth and development in the learning analytics community. One of the developments to recently emerge as a recognized special interest group in Learning Analytics is the sub-field of Multimodal Learning Analytics (MmLA). In this paper we consider a future trajectory for MmLA that intersects with the cross-cutting 21st century skill of collaboration. Teaching collaboration is seldom the focus of formal, or informal learning experiences, as students and teachers rarely receive feedback on their collaboration process. Instead, feedback is normally reduced to an outcome measure, or requires a level of human analysis that is intractable at scale. We see a unique opportunity for MmLA to promote collaboration literacy, and for collaboration literacy to be a common space in which to grow MmLA. Concretely, MmLA can provide the theoretical and technological innovations needed to create tools that support the evaluation, assessment and development of collaborative skills. As a first step in this direction, this paper presents a framework for collaboration literacy that consists of four levels of increasing complexity. We describe examples of current work in the first three levels of the framework, and situate the fourth level as an aspirational goal for the field of MmLA. We also discuss some of the key challenges that need to be solved to facilitate increased adoption of a collaboration literacy feedback tool, and MmLA more broadly. Ultimately, we argue that the development of such a tool could be instrumental in introducing new ways for building collaboration literacy.

Keywords: multimodal feedback, data fusion, framework, data capture

1 INTRODUCTION

Being able to effectively practice collaborative problem solving has been widely identified as one of the key skills that is needed to succeed, learn, and work in the 21st-century (Dede, 2010). Despite its importance, the development of collaboration skills usually consists of exposing students to a series of collaborative experiences with limited scaffolding. Furthermore, most of the feedback is about the final product of the collaboration, with almost no feedback about student in-situ collaboration skills (Lai, 2011). These “activity-as-instruction” approaches for the development of collaboration skills usually overlooks individual performance (Lai, 2011) and severely restrict the potential learning opportunities from otherwise carefully designed activities.

This focus on evaluation of collaboration artifacts instead of evaluation of the collaborative process is not an oversight or pedagogically-justified preference. It is a predictable result of both a lack of literacy and capacity to provide such feedback. It is easier to provide and receive feedback about a concrete artifact that has a predefined physical or digital form and can be easily shared between participants, than to give feedback about a sequence of remembered actions that are not always shared (or remembered) by students or teachers. While analyzing the collaboration process is possible, and it is routinely done for research purposes (e.g.: Kleinsmann, Deken, Dong, and Lauche (2012); Berland, Davis, and Smith (2015); Shaffer, Collier, and Ruis (2016)), the level of observation, coding and analysis currently required is scarcely practical for a teacher (or students) during a normal collaborative activity. Providing feedback about the collaboration behaviour of each student is a desirable goal. However, the current practice is too laborious and time consuming to be routinely used.

There have been several proposals on how to improve the feasibility of teaching collaboration skills in regular educational contexts. These proposals could be summarized in what Griffin and Care (2014) note when talking about how to improve the teaching and assessment of 21st century skills: “New forms of data collection needed to be devised, and methods of analysing those new forms of data need to be identified and tested”. This challenge perfectly aligns with the goal of Multimodal Learning Analytics (Ochoa & Worsley, 2016) (MmLA): capturing, analyzing and fusing several streams of data to better understand and improve learning processes. MmLA is uniquely positioned to bridge the gap between what is a desirable pedagogical approach (providing detailed feedback about student collaboration practices) and what is practical (a pedagogical/technical tool that can be easily used in general collaborative activities in the classroom). MmLA provides the technical tools to easily capture human behaviour using low-cost, synchronized, multimodal sensors (Domínguez, Chiluíza, Echeverría, & Ochoa, 2015; Martínez-Maldonado, Echeverría, Santos, Santos, & Yacef, 2018) and to estimate learning-relevant constructs based on the analysis and fusion of that multimodal data (Di Mitri, Schneider, Specht, & Drachsler, 2018; Worsley, 2018a)

As we consider the design of learning analytics tools that can facilitate the development of collaboration, we argue that a primary focus should be to rapidly provide feedback that is clear and actionable. To do this, it should be able to fulfil the following requirements:

- It should be able to automatically provide a detailed multimodal transcript (Ochoa, Chiluíza, et al., 2018) that summarizes the relevant actions that occurred during the collaboration activity. This information can support reflection by the teacher or the students. Teacher-facing dashboard for students collaboratively building database diagrams (Granda, Echeverría, Chiluíza, & Wong-Villacrés, 2015) and an Emergency Room behaviour reflection tool for nurses in training (Martínez-Maldonado et al., 2019) are examples that satisfy this requirement.
- It should also provide estimates of collaboration-relevant constructs. These estimates should be based on objectively measured quantities that augment an individual's capability to understand their collaboration behaviour. An example in another domain is the analytical report provided by an oral presentation feedback tool (Ochoa, Domínguez, et al., 2018).

- It should integrate seamlessly into the collaboration activity. For example, an instrumented table that displays the percentage of conversation-time used by each participant around the table (Bachour, Kaplan, & Dillenbourg, 2010) would satisfy this type of requirement.
- It should support different types of classroom orchestration (Dillenbourg, Prieto, & Olsen, 2018) and not be designed for just one type of collaborative activity. Alternatively, it should have utility for different learning context.

This list of requirements currently lies in the fuzzy frontier of what is pedagogically beneficial and technically possible. The main contribution of this paper is detailing the guidelines and a possible road-map to build a tool that is both theoretically-grounded and technologically feasible. Section 2 will introduce the concept of collaboration literacy and a novel multi-level framework to connect collaboration constructs with existing MmLA research. Section 3 will present the technical challenges to convert the framework into a functioning tool that could be easily deployed in real-world scenarios. Finally, we conclude with remarks on a potential path forward for MmLA and collaboration literacy.

2 DEFINING COLLABORATION LITERACY

We conceptualize collaboration literacy as the ability to ascertain and respond to changes in the quality of a collaborative experience. From the student perspective this amounts to being conscious of one's own contribution to a group, as well as the awareness and ability to intervene in order to ensure a strong collaboration. From the teacher perspective this includes awareness of how different groups are progressing, being able to respond to those groups in a timely fashion, and developing prompts and activities that afford good collaboration.

It is easy to adopt a simplistic perspective around collaboration quality that consists of generic labels for “good” and “bad” collaboration. Without question, there are practices that can promote more or less effective collaboration. At the same time, however, there are a number of more complex practices and behaviours that differentially contribute to the nature of a collaboration. Within this paper, we provide a framework for thinking about the complexity of different constructs. We position the framework as being important to guiding on-going research at the intersection of MmLA and collaboration. Concretely, it provides a framework that researchers can use to position their work, and also provides aspirational goals for where their work might go. Table 1 identifies collaboration related constructs, the research that supports their salience and where we place each construct relative to the multi-level system that we describe below. This list is not exhaustive. Advancing collaboration literacy helps participants learn to recognize the different levels at which one might quantify collaboration quality.

Note: Though we use literal versus semantic in thinking about the different levels, these are approximations, as these terms are not true binaries. Instead, there is a continuum between literal and semantic that is difficult to represent in a strict set of four categories. Nonetheless, we make every effort to draw clear distinctions between the different levels of the framework, and propose that multimodal data can afford a more semantic representation of a student's actions or perceived state.

Table 1. Levels of collaboration analysis, their associated constructs and example references

Level	Constructs	Example Reference
1	Text Contributions	Leeuwen van Leeuwen, Janssen, Erkens, and Brekelmans (2015)

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

Unimodal Literal Individual	Head Position	Worsley, Scherer, Morency, and Blikstein (2015)
	Affective State	Richey, D'Angelo, Alozie, Bratt, and Shriberg (2016)
	Tactile Engagement	Worsley (2018a)
2 Unimodal Literal Multi-party	Entrainment	Lubold and Pon-Barry (2014)
	Body Synchrony	Cukurova, Luckin, Millán, and Mavrikis (2018)
	Joint Visual Attention	B. Schneider et al. (2018)
	Turn Taking	Devault, Mell, and Gratch (2015)
3 Multimodal Semantic Individual	Turn Management	Emma M. Mercier, Higgins, and da Costa (2014)
	Questioning	Hmelo-Silver and Barrows (2008)
	Monitoring	Hmelo-Silver and Barrows (2008)
	Summarizing	Wise and Chiu (2011)
4 Multimodal Semantic Multiparty	Rapport	Gratch, Wang, Gerten, Fast, and Duffy (2007)
	Negotiating Interactive	Emma Mary Mercier, Higgins, Burd, and Joyce-Gibbons (2012)
	Convergent Conceptual Change	Roschelle (1992)
	Struggling	Bassiou et al. (2016)

2.1 Level 1: Unimodal, Literal, Individual

The first level for examining collaboration quality involves looking at how a given individual is contributing or engaging through a single modality. Within research on collaboration, the modalities of speech and text tend to be privileged, as verbal, or textual, engagement is often a precursor to an effective collaboration (Hmelo-Silver & Barrows, 2008; Emma M. Mercier et al., 2014; Richey et al., 2016; van Leeuwen et al., 2015). Within this paradigm, researchers might look at the quantity and quality of speech and/or text generation. For example, researchers might look at the speech fraction, or the amount of time someone is talking, relative to the total time of the activity. Speech fraction values approaching 0 or 1 are likely to be indicative of interactions that were not very collaborative. Text also tends to be a rich modality for gleaning insights about the nature of an interaction. With the assistance of computational tools (i.e., Lightside (Mayfield & Rosé, 2013) or Natural Language Toolkit (Loper & Bird, 2002)) or through human annotation, researchers can begin to develop a better understanding of an individual's cognitive, social, or emotional state during a given collaborative activity. Inferences about participant state can also be inferred from the use of video or voice technology. These unimodal data points can be informative for characterizing the relative success of the interaction, and for enabling easy identification of noticeable changes in individual participation. Early work in MmLA demonstrated how looking at a single modality, could help predict collaboration among students completing math problems. Specifically, Ochoa et al. (2013) found that using various simple features for how fast someone writes, or draws, the percentage of time they use the calculator and how much they mention numbers or mathematical terms, are good proxies for predicting their level of expertise within a group. Hence, in certain situations, a unimodal, individual approach can provide a reasonable starting point for ascertaining the nature of a group collaboration.

2.2 Level 2: Unimodal, Literal, Multi-party

Level 1 looked at how a given individual may be engaging with a specific task through a single modality. Level 2 extends these measures to being multi-party. This has been a primary paradigm utilized by collaboration researchers (Hmelo-Silver & Barrows, 2008; Emma Mary Mercier et al., 2012; Emma M. Mercier et al., 2014; Roschelle, 1992; Wise & Chiu, 2011). For example, in the case of verbal contribution, instead of looking at how much someone talked, researchers consider the nature of turn-taking within a given group. In addition to looking at how speaking turns are distributed across a group, researchers might take a more qualitative approach, and code participant utterances for ways

that turns are managed, and the ways that a given speaker's idea is taken up by the other participants. For instance, researchers might label when someone is responding to a previous utterance, or examine when a given utterance signals student agreement with a given idea (Richey et al., 2016; Wise & Chiu, 2011). In these cases, a given utterance only becomes relevant in context of the surrounding utterances, and in the context of the other individuals within the space. This is one of the important pieces added by considering multi-party collaboration. The multi-party level also allows for more direct consideration of the power relations or social dynamics of a given group. Cukurova et al. (2018) provide an informative analysis using level 2 collaboration to study social dynamics among a group of students completing a hands-on task. Specifically, they analyze hand gesticulation to look at the extent of body synchrony among participants. They also answer questions about whether or not group participants appear to be exerting the same amount of body movement at a given time, or if the amount of physical movement is unevenly distributed. Another good example of unimodal, multiparty collaboration is visual joint attention (B. Schneider et al., 2018). Joint visual attention refers to instances where two, or more, individuals are looking at the same location, or object, at, roughly, the same time. Many prior studies have highlighted the importance of joint visual attention for promoting learning and perceptions of collaboration quality. It can also be indicative of power relations, when considering who, within a given collaborative group, receives the visual attention of their peers when speaking. Broadly speaking, Level 2 moves us closer to multi-party metrics, and, adds additional challenges and opportunities. Data must be synchronized across different participants, but upon doing so, it becomes easier to identify group-level patterns that emerge.

2.3 Level 3: Multimodal, Semantic, Individual

Level 3 introduces semantics and multimodality. Whereas Level 1 analysis of text using tools like Lightside (Mayfield & Rosé, 2013) and Coh-metrix (Graesser, McNamara, & Kulikowich, 2011), provide some degree of utterance understanding and intent interpretation, level 3 surfaces the opportunities for using data from different modalities to represent perceived student state. For instance, we can represent user engagement based on the presence or absence of speech. If, though, a user is generating speech without addressing their peers, as detected through head pose estimation, or eye gaze, it becomes less likely that the user is engaging in the collaborative task. Returning to the example of verbal contributions, when we talk about semantics we are examining the words that are uttered, and deriving meaning from those collections of words. Taking a multimodal and semantic perspective with gesticulations refers to identifying specific gestures in the gesticulations that a student is making, and, perhaps, connecting those gestures with user spoken utterances. A simple gesture one might make in the context of a classroom is raising one's hand, or pointing. In order to ascertain these gestures, one has to rely on a semantic understanding of the student's gesticulation. In the case of a student pointing, there is an additional multimodal component of understanding what they are pointing to. Prior work in learning analytics has begun to consider collaboration at this level. For example, Worsley and Blikstein (2018), inspired by Scherr and Hammer (2009), develop multimodal representations of individuals in collaborative pairs. Scherr and Hammer (2009) describes epistemological frames, which constitute a combination of modalities (e.g., speech, head pose, gesticulation) that, when combined, provide a sense of the type of activity a student is undertaking at a given time. Each of the epistemological frames: Discussion, Lecture, Teaching Assistant and Joking; is characterized by a different combination of the aforementioned modalities. Worsley and Blikstein (2018) extend this idea by using an electro-dermal activation, speech and hand/wrist movement to

identify four representative modes of collaboration during a hands-on building task. Importantly, that particular analysis was primarily done on an individual basis and did not consider the ways individuals reacted to one another, which is a key differentiator between Level 3 and Level 4.

2.4 Level 4: Multimodal, Semantic, Multi-party

Level 4 elevates the level 3 measures to multi-party inferences. Here, we consider measures like shared understanding and convergent conceptual change. Roschelle's work (1992) on convergent conceptual change highlights ways that groups negotiate the collaborative learning process through a combination of gestures and spoken turns. Specifically, convergent conceptual change is

[C]haracterized by: (a) the production of a deep-featured situation, in relation to (b) the interplay of physical metaphors, through the constructive use of (c) interactive cycles of conversational turn-taking, constrained by (d) the application of progressively higher standards of evidence for convergence.

Convergent conceptual change represents a complex interplay of student actions around a shared task. The component constituents of the interaction can be reasonably characterized through semantic, multimodal interpretation of gestures and verbal utterances. For example, speaking turns can be labeled through speech recognition, and physical gesticulations analyzed for specific gestures. The semantics of student utterances can be interpreted for different measures of cohesion, or argumentation, and combined with the corresponding gestures. However, the actual demonstration of conceptual change requires an additional level of inference that goes beyond the individual. It necessitates that an individual's data be interpreted relative to the data of the other participants.

Broadly speaking, Level 4 measures require a semantic and multimodal interpretation of group behavior, often across time and at variable time scales. This is an area of research that has received little attention from the MmLA community, and reasonably involves the highest amount of complexity. It also requires a certain level of accuracy within the level 3 measures and data representations. Part of what we propose in this paper is that developing theories and representations of collaboration that mirror the complexity of convergent conceptual change, is one of the opportunities for the future of MmLA.

We argue that all of the levels could benefit from MmLA. Levels 1 and 2, are forms of interaction that can be reasonably approximated through current artificial intelligence technologies. Levels 3 and 4, could, at present, be researched through a combination of human-machine analyses, with the eventual goal of being incorporated into real-time tools. In consideration of these factors, the section to follow describes technical challenges that we are exploring to realize developing such a tool.

3 TECHNICAL CHALLENGES

Most of the state-of-the-art in capturing and analyzing collaboration construct is the result of lab-based prototypes in which instructors and learners are only involved during the data-capturing phase. Building a tool that can be used on a regular basis in common learning contexts to improve collaboration literacy needs not only models to convert raw recording data into high-level constructs, but also a technical infrastructure that makes it deployable, scalable and acceptable. This section will provide a discussion of some challenges and potential solutions.

3.1 Type of Sensors and Modalities

There is large range of sensors and modalities that have been used in MmLA studies (Di Mitri et al., 2018; Ochoa, 2017). There is an inherent tension between the desire to capture as many modalities as possible and the complexity and intrusiveness of the recording apparatus. Given the set of constructs defined in the previous section, the recommended trade-off between the two extremes is a combination of high-definition horizontal 360 degrees video (captured with a simple camera and a fisheye lens) and directional audio. Apart from existing log data captured by digital systems, video cameras and microphones have been the sensors of choice in MmLA research. This preference for audio and video is because they can reliably capture the primary forms of human communication, have high information density, align with the information captured by human senses, are low cost, are easy to deploy and are non-intrusive (Worsley, 2018b). Different from considerable prior work, however, is the inclusion of microphone arrays, which allow for the collection of directional audio, and 360 degree cameras, which provide for substantial coverage of a given learning environment. Based on current state-of-the-art in MmLA, these sensors enable the capture of posture, gaze, facial expression, hand gestures and actions, position, speaker identity, speech verbal and non-verbal features (Ochoa, 2017). While other sensors (e.g., biophysiological sensors) are available for instrumenting people and learning environments, we want to be careful about balancing the utility of the sensors, with concerns about data privacy and ethics and deployability.

3.2 Synchronicity

Synchronicity of the recordings allow the fusion of information from different modalities. Synchronization precision depends on the type of signals being combined and the type of analysis to be conducted on the resulting features. To establish the level of synchronization needed, we reference Newell's time scale of human actions (Newell, 1994). Newell defined different time spans for several learning-related human actions and reactions. These speeds are divided into several bands according to the type of process that generates it. The bands are biological (100 microseconds to 10 milliseconds), cognitive (100 milliseconds to 10 seconds), rational (1 minute to hours) and social (days to months). The most relevant aspects of human collaboration, and also the ones that are deliberate by the student and perceptible to a human observer, are in the cognitive, rational and social bands. The lower bound for these kinds of signals is a tenth of a second. This level of synchronization is perfectly achievable with current off-the-shelf technologies. For example, the Social Signal Interpretation (SSI) framework (Wagner et al., 2013), which allows for synchronization on the order of milliseconds even when the recording is distributed across different devices, provides a viable solution for simplifying synchronization.

3.3 Deployment

Having a recording apparatus that can be setup and operated by non-experts is a main challenge in moving from lab conditions to real learning settings. Two options have been tested to facilitate this change: fixed pre-configured setups and user-friendly mobile setups. In the first option, a complex recording system is built and configured ahead of time. With this system, users are only permitted to complete two actions: turning on the recording, and turning off the recording. This strategy has been followed by Ochoa, Domínguez, et al. (2018) in their widely deployed Oral Presentation Feedback system. A commercial example of this type of devices is the Meeting Owl, a videoconference camera

for group meetings. The creation of these kinds of mobile devices requires both engineering and user-based-design efforts to create easy-to-use interfaces for minimalist hardware. Given that collaboration activities could happen in any classroom, it is recommended that systems for collaboration literacy feedback follow this second approach where the recording device is mobile and easily operable by the collaboration activity participants.

3.4 Real-time vs. Post-hoc Feedback

When the feedback is provided can have an important effect on its usefulness and actionability. The difference is exemplified by two types of multimodal oral presentation feedback tools. The first one, introduced by (J. Schneider, Börner, Rosmalen, & Specht, 2015) presents simple feedback about posture, gaze and volume in real-time to the presenter through an augmented reality visor. The second, a system proposed by (Ochoa, Domínguez, et al., 2018) provides a more detailed feedback about the same modalities through a multimodal report but only after the presentation is finished. Both system show positive learning gains. It is not clear what system is more appropriate to develop presentation skills, or if a combination of the two is the right answer. A system for collaboration literacy feedback should explore both real-time multimodal signals and post-hoc reflection reports in order to find which one has a stronger impact on learning different collaboration literacy constructs. It is also important to consider the computational requirements that real-time feedback have in the multimodal extraction and fusion component.

3.5 Individual vs. Group Feedback

Feedback can be provided privately to individuals about their individual collaboration behaviour. However, there is also an element of group dynamics that cannot be explained by individual contributions alone. Some collaboration constructs only make sense at the group level. Also, exposing individual feedback to the group has the potential to violate the right of privacy of the individual. Studies such as (Archer-Kath, Johnson, & Johnson, 1994) where the individual vs. group feedback is empirically tested should be conducted to test impact of collaboration literacy feedback interfaces.

3.6 Automated vs. Human-augmented Feedback

Even with current advances in artificial intelligence, there are certain aspects of collaboration behaviour that cannot be detected or processed by current automated systems. Human feedback has the potential to be of higher quality than automated systems. However, human feedback is not without limitations. Most importantly, it is not scalable, and is subject to bias. Combining the right proportion of both types of feedback seems to be the right approach. This determination is also an open research question that should be addressed after determining which types of collaboration constructs can be accurately and reliably estimated automatically, and which ones still need human input. Moreover, systems for collaboration literacy feedback could provide an interface for human instructors to focus their capabilities on resolving difficult-to-judge cases or constructs. For example, an automated system could provide feedback about turn management and questioning, while an annotated multimodal transcript (such as in (Echeverria et al., 2018)) could be provided to instructors to focus their attention on key moments of the collaborative activity. The combination of automated and human-augmented feedback could also support teachers, students and researchers focusing on higher level collaboration constructs.

4 CONCLUSIONS

In this paper, we argue that MmLA is uniquely poised to analyze collaborative learning environments. Moreover, we propose a framework for considering the different levels of complexity of collaborative problem solving, with the goal of supporting the development of collaboration literacy, a form of literacy that receives little formal attention within mainstream, and even progressive learning experiences. Enacting the creation of collaboration literacy feedback tools can potentially be achieved through a combination of low-cost audio and video data capture technology, in conjunction with the development of robust multimodal fusion, and multimodal feedback strategies. Determining the design of collaboration literacy feedback tools will involve research and development along several of the dimensions outlined in this paper, and likely some additional dimensions that have yet to be identified. Nonetheless, we position the ideas included in this paper as a concrete, constructive and feasible research agenda for simultaneously advancing MmLA and collaboration literacy. Levels 1 through 3 of the framework represent constructs that MmLA can address in the short-term. These constructs can be enacted through unimodal and multimodal features that are available through current artificial intelligence technologies. Level 4 represents an aspirational goal for MmLA. Such investigations have the opportunity to drive new theories and conjectures about the complexities of group collaboration, much like Roschelle's work (1992) on convergent conceptual change. Finally, this work, as a whole suggests the need to close the gap in MmLA by promoting important, real-time feedback (Bassiou et al., 2016) and to carefully consider issues of ethics and data privacy.

Our hope is that this paper will help provide direction for the field to more quickly converge towards the development of common apparatus for distributed data collection, shared measures, and consistent feature extraction and fusion algorithms.

REFERENCES

- Archer-Kath, J., Johnson, D. W., & Johnson, R. T. (1994). Individual versus Group Feedback in Cooperative Groups. *The Journal of Social Psychology*, 134(5), 681-694. doi:10.1080/00224545.1994.9922999
- Bachour, K., Kaplan, F., & Dillenbourg, P. (2010). An Interactive Table for Supporting Participation Balance in Face-to-Face Collaborative Learning. *IEEE Trans. Learn. Technol.*, 3(3), 203-213. doi:10.1109/tlt.2010.18
- Bassiou, N., Tsiartas, A., Smith, J., Bratt, H., Richey, C., Shriberg, E., . . . Alozie, N. (2016). *Privacy-Preserving Speech Analytics for Automatic Assessment of Student Collaboration*. Paper presented at the Interspeech 2016.
- Berland, M., Davis, D., & Smith, C. P. (2015). AMOEBA: Designing for collaboration in computer science classrooms through live learning analytics. *International Journal of Computer-Supported Collaborative Learning*, 10(4), 425-447.
- Cukurova, M., Luckin, R., Millán, E., & Mavrikis, M. (2018). The NISPI framework: Analysing collaborative problem-solving from students' physical interactions. *Computers & Education*, 116, 93-109.
- Dede, C. (2010). Comparing frameworks for 21st century skills. *21st century skills: Rethinking how students learn*, 20, 51-76.
- Devault, D., Mell, J., & Gratch, J. (2015). *Toward Natural Turn-Taking in a Virtual Human Negotiation Agent*. Paper presented at the AAAI Spring Symposium on Turn-Taking and Coordination in Human-Machine Interaction.

- Di Mitri, D., Schneider, J., Specht, M., & Drachsler, H. (2018). From signals to knowledge: A conceptual model for multimodal learning analytics. *Journal of Computer Assisted Learning*, 34(4), 338-349.
- Dillenbourg, P., Prieto, L. P., & Olsen, J. K. (2018). Classroom Orchestration. In *International Handbook of the Learning Sciences* (pp. 180-190): Routledge.
- Domínguez, F., Chiluíza, K., Echeverría, V., & Ochoa, X. (2015). *Multimodal selfies: Designing a multimodal recording device for students in traditional classrooms*. Paper presented at the Proceedings of the 2015 ACM on International Conference on Multimodal Interaction.
- Echeverría, V., Martínez-Maldonado, R., Granda, R., Chiluíza, K., Conati, C., & Shum, S. B. (2018). *Driving data storytelling from learning design*. Paper presented at the Proceedings of the 8th International Conference on Learning Analytics and Knowledge.
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational researcher*, 40(5), 223-234.
- Granda, R. X., Echeverría, V., Chiluíza, K., & Wong-Villacrés, M. (2015). *Supporting the assessment of collaborative design activities in multi-tabletop classrooms*. Paper presented at the 2015 Asia-Pacific Conference on Computer Aided System Engineering.
- Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. (2007). *Creating Rapport with Virtual Agents*, Berlin, Heidelberg.
- Griffin, P., & Care, E. (2014). *Assessment and teaching of 21st century skills: Methods and approach*: Springer.
- Hmelo-Silver, C. E., & Barrows, H. S. (2008). Facilitating Collaborative Knowledge Building. *Cognition and Instruction*, 26(1), 48-94. doi:10.1080/07370000701798495
- Kleinsmann, M., Deken, F., Dong, A., & Lauche, K. (2012). Development of design collaboration skills. *Journal of Engineering Design*, 23(7), 485-506. doi:10.1080/09544828.2011.619499
- Lai, E. R. (2011). *Collaboration: A literature review*. Retrieved from
- Loper, E., & Bird, S. (2002). NLTK: the natural language toolkit. *arXiv preprint cs/0205028*.
- Lubold, N., & Pon-Barry, H. (2014). *Acoustic-Prosodic Entrainment and Rapport in Collaborative Learning Dialogues*. Paper presented at the Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge, Istanbul, Turkey.
- Martínez-Maldonado, R., Echeverría, V., Elliott, D., Axisa, C., Power, T., & Shum, S. B. (2019). Making the Design of CSCL Analytics Interfaces a Co-design Process: the Case of Multimodal Teamwork in Healthcare.
- Martínez-Maldonado, R., Echeverría, V., Santos, O. C., Santos, A. D. P. D., & Yacef, K. (2018). *Physical learning analytics: A multimodal perspective*. Paper presented at the Proceedings of the 8th International Conference on Learning Analytics and Knowledge.
- Mayfield, E., & Rosé, C. P. (2013). LightSIDE: Open source machine learning for text. In *Handbook of Automated Essay Evaluation* (pp. 146-157): Routledge.
- Mercier, E. M., Higgins, S., Burd, E., & Joyce-Gibbons, A. (2012). *Multi-touch technology to support multiple levels of collaborative learning in the classroom*.
- Mercier, E. M., Higgins, S. E., & da Costa, L. (2014). Different leaders: Emergent organizational and intellectual leadership in children's collaborative learning groups. *International Journal of Computer-Supported Collaborative Learning*, 9(4), 397-432. doi:10.1007/s11412-014-9201-z
- Newell, A. (1994). *Unified theories of cognition*: Harvard University Press.
- Ochoa, X. (2017). Multimodal learning analytics. In G. S. Charles LANG, Alyssa WISE, and Dragan GAŠEVIĆ (Ed.), *The Handbook of Learning Analytics* (Vol. 1, pp. 129-141).
- Ochoa, X., Chiluíza, K., Granda, R., Falcones, G., Castells, J., & Guamán, B. (2018). *Multimodal Transcript of Face-to-Face Group-Work Activity Around Interactive Tabletops*. Paper presented at the CrossMMLA@ LAK.
- Ochoa, X., Chiluíza, K., Méndez, G., Luzardo, G., Guamán, B., & Castells, J. (2013). *Expertise estimation based on simple multimodal features*. Paper presented at the Proceedings of the 15th ACM on International conference on multimodal interaction.

- Ochoa, X., Domínguez, F., Guamán, B., Maya, R., Falcones, G., & Castells, J. (2018). *The rap system: automatic feedback of oral presentation skills using multimodal analysis and low-cost sensors*. Paper presented at the Proceedings of the 8th international conference on learning analytics and knowledge.
- Ochoa, X., & Worsley, M. (2016). Augmenting Learning Analytics with Multimodal Sensory Data. *Journal of Learning Analytics*, 3(2), 213-219.
- Richey, C., D'Angelo, C., Alozie, N., Bratt, H., & Shriberg, E. (2016). *The SRI Speech-Based Collaborative Learning Corpus*. Retrieved from
- Roschelle, J. (1992). Learning by Collaborating: Convergent Conceptual Change. *The Journal of the Learning Sciences*, 2(3), 235-276. Retrieved from <http://www.jstor.org/stable/1466609>
- Scherr, R. E., & Hammer, D. (2009). Student Behavior and Epistemological Framing: Examples from Collaborative Active-Learning Activities in Physics. *Cognition and Instruction*, 27(2), 147-174. doi:10.1080/07370000902797379
- Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2018). Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups. *International Journal of Computer-Supported Collaborative Learning*, 13(3), 241-261.
- Schneider, J., Börner, D., Rosmalen, P. v., & Specht, M. (2015). *Presentation Trainer, your Public Speaking Multimodal Coach*. Paper presented at the Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, Seattle, Washington, USA.
- Shaffer, D. W., Collier, W., & Ruis, A. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3(3), 9-45.
- van Leeuwen, A., Janssen, J., Erkens, G., & Brekelmans, M. (2015). Teacher regulation of cognitive activities during student collaboration: Effects of learning analytics. *Computers & Education*, 90, 80-94. doi:<https://doi.org/10.1016/j.compedu.2015.09.006>
- Wagner, J., Lingensfelder, F., Baur, T., Damian, I., Kistler, F., Andr, E., & #233. (2013). *The social signal interpretation (SSI) framework: multimodal signal processing and recognition in real-time*. Paper presented at the Proceedings of the 21st ACM international conference on Multimedia, Barcelona, Spain.
- Wise, A. F., & Chiu, M. M. (2011). Analyzing temporal patterns of knowledge construction in a role-based online discussion. *International Journal of Computer-Supported Collaborative Learning*, 6(3), 445-470. doi:10.1007/s11412-011-9120-1
- Worsley, M. (2018a). *(Dis) engagement matters: Identifying efficacious learning practices with multimodal learning analytics*. Paper presented at the Proceedings of the 8th International Conference on Learning Analytics and Knowledge.
- Worsley, M. (2018b, 1/1). *Multimodal learning analytics' past, present, and, potential futures*. Paper presented at the CEUR Workshop Proceedings.
- Worsley, M., & Blikstein, P. (2018). A Multimodal Analysis of Making. *International Journal of Artificial Intelligence in Education*, 28(3), 385-419. doi:10.1007/s40593-017-0160-1
- Worsley, M., Scherer, S., Morency, L.-P., & Blikstein, P. (2015). *Exploring Behavior Representation for Learning Analytics*. Paper presented at the Proceedings of the 2015 ACM on International Conference on Multimodal Interaction, Seattle, Washington, USA.

SPARK: A Learning Analytics Leadership Framework

Shane Dawson

University of South Australia
shane.dawson@unisa.edu.au

Abelardo Pardo

University of South Australia
abelardo.pardo@unisa.edu.au

George Siemens

University of South Australia
gsiemens@gmail.com

ABSTRACT: Learning analytics has developed rapidly over the past decade, providing extensive contributions to the science of learning. For continued success and impact, LA must move from the research lab and into organizational practices. Innovative insights must scale to innovative practices. In order to make this transition, an exploration of the role of leadership is required. This workshop introduces SPARK – a framework for leadership to deploy and evaluate the impact of LA. This framework accounts for the complexities of learning environments, organizational policies, and external environment practices while guiding leadership teams to deploy LA initiatives scalable to achieve organizational impact.

Keywords: Learning Analytics, Complexity Leadership, Higher Education, Technology Adoption

1 INTRODUCTION

Learning Analytics (LA) juxtaposes multiple strands of research such as computer science, education, psychology and statistics to leverage their individual theoretical and methodological strengths to bring about novel insights to the learning process. Such combinatorial work can and has generated new and significant research (Mangaroska & Giannakos, 2018). However, in spite of the progress made in LA over the past decade, the practicality of deploying the research outcomes at scale have remained problematic. This is, in part, due to the varying contexts in which such socio-technical research is conducted. There is a critical gap in LA research and development related to the lack of large scale and mature instantiations in education settings (Dawson, Joksimovic, Poquet, & Siemens, 2019; Ferguson et al., 2014).

2 COMPLEXITY LEADERSHIP

Education is a complex adaptive system (CAS). Such systems are characterized as containing numerous interacting agents that give rise to new and emergent self-organizing behaviors (Siemens, Dawson, & Eshleman, 2018). Thus, for education leaders enacting change is seldom successful through rigid structures of governance. Instead, leaders must learn to adapt and identify the key actors and points of leverage to proactively influence organizational networks and guide actions and responses towards more intended goals. The complexities associated with developing LA infrastructure, staff skills and

training are not new and have been explored in the literature (For eg: Arnold, Lonn, & Pistilli, 2014; Colvin et al., 2015; Greller & Drachsler, 2012). For LA projects there are a range of factors including privacy and ethics, data collection, analysis, sensemaking, institutional strategy, and capacity building that can impede adoption and derail progress (Tsai, Poquet, Gašević, Dawson, & Pardo, 2019). Consequently, many education institutions partner with corporate providers to “fast track” organizational data gathering for analytics and reporting. However, while the technical infrastructure for LA may be present there is a dearth of evidence illustrating how such analytics are actually improving and informing practice. The availability of data and a developed technical infrastructure are merely first steps in progressing towards a data-informed institution.

The ability to transition LA outcomes from the fringes of educational research to mainstream practice requires an interplay between two interlocked systems: the technical and social-contextual. To date, greater emphasis has been devoted to developing the technical system associated with LA (Dawson et al., 2018). In contrast, the examination of the social aspects of how learning analytics are applied and adopted have been comparatively limited. This workshop approaches the challenge of adopting LA to improve teaching and learning by examining the role of leadership in navigating these overarching systems. We propose SPARK: a leadership framework based in complexity leadership theory (CLT) (Uhl-Bien, Marion, & McKelvey, 2007), to plan, monitor, and develop the technical and socio-contextual aspects of LA adoption. The SPARK framework explicitly characterizes how LA leaders navigate the *productive tensions* necessary to scale LA while accounting for the growing internal and external complexity of modern universities.

3 SPARK FRAMEWORK

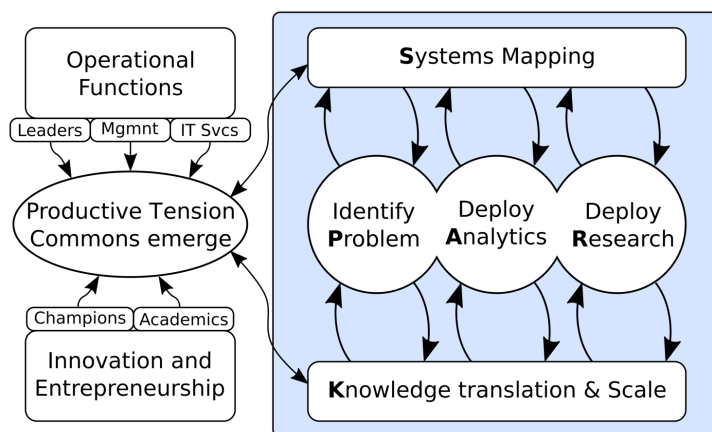


Figure 1 SPARK framework

Numerous authors have proposed multiple approaches for LA adoption (e.g. Ferguson et al., 2014). However, these models are conceptual and lack sufficient details for LA leaders to translate outcomes to more scaled and operationalized endeavors. To address this issue, we propose the SPARK framework (Figure 1). The framework comprises 5 inter-related dimensions – System mapping; Problem identification; Analytics; Research and Knowledge translation and scale.

3.1 Systems

The first dimension relates to mapping the system to unpack the actors spanning the operational (technical) and entrepreneurial (social) systems. For LA projects this will involve technical and support staff (operational) as well as teachers and researchers (entrepreneurial). As SPARK is concerned with the productive tensions between the operational and entrepreneurial functions, systems mapping identifies the processes that enable implementation to establish allies and progression towards potential networks of influence. Building a base of potential capital or organizational power to

maintain momentum for the project as translation to mainstream practice will disrupt other components in the operational system. Knowing who is involved, their roles and responsibilities is more essential than knowing the discrete technicalities such as data queries and storage. In this phase leaders establish the working context, culture and communication channels for the project.

3.2 PROBLEM IDENTIFICATION

Problem identification serves to align the project with the institution's strategic goals. This step entails developing the project goals and longer-term vision. It is important to note that as the outcomes of one stream of the LA implementation are developing, several external project areas may also be positively or negatively impacted. Systems integration and thinking are critical in this stage. Leadership teams need to sort through conflicts and tensions to establish next steps and progress. That is – to return to the site of productive tensions to resolve project conflicts and directions. The problem identification serves to align all stakeholders with the central issues to address, to ensure consistency in the project team with regards to goals and resources, and to propose an initial minimal viable product to garner evidence for future support and persuasion.

3.3 ANALYTICS

This dimension builds on problem identification and serves to ensure all stakeholders are in agreement regarding the measures and targets that will indicate success. The establishment of identified targets and analytics ensures that findings drawn from the Research phase are sufficient to first demonstrate progress in addressing the noted problem; and secondly to establish the necessary data and evidence to influence change in the organization. Leaders in this stage need to assure that the research processes have the right rigor and alignment to support decision-making processes, as well as addressing core processes such as data infrastructure and data access.

3.4 RESEARCH

This phase involves the investigation and evaluation of pilot studies formed through the prior phases of the SPARK framework. The goal of the research is to empirically demonstrate impact. This requires communication with the key actors identified in the Systems mapping. Early discussion of impact and changes to operational systems can commence. Demonstration of impact to influence senior staff and key actors in the organization is necessary to provide support for more advanced pilots. The outcomes are used to identify the next steps to progress from pilot phase to organizational transformation.

3.5 KNOWLEDGE BROKERING AND TRANSITION TO SCALE

This phase requires building multiple sites of strong ties to aid diffusion of the innovation leading to a change in teaching and learning behavior. Hence, from the pilot study individuals with strong relationships with peers must act as the first point for diffusion. This may relate to a course coordinator working with program peers and other discipline-based teachers. This knowledge brokering role requires multiple individuals with direct experience with the LA pilot and outcomes to “broker” and engage peers in the benefits, limitations and future opportunities for change and engagement. At all times there is a strong requirement for support and professional development.

4 ORGANIZATIONAL DETAILS

The proposed structure for the half-day, open workshop is:

- Description of the context and the proposed SPARK framework (30 mins)
- Interactive session to unpack the elements required in each dimension and how these connect with the participant's context. (30 mins)
- Case studies: Two case studies are shown, (successful and unsuccessful). Groups will diagnose problems and propose changes to promote institutional uptake and wrap up (1.5 hour)

By the end of the workshop the participants should be able to:

- Appraise the level of adoption of LA in their institutions
- Determine the required relationships and actions to support this adoption
- Design and recommend a strategy for institutional adoption of LA
- The workshop will be disseminated through a web site including details about the framework.

REFERENCES

- Arnold, K. E., Lonn, S., & Pistilli, M. D. (2014). *An Exercise in Institutional Reflection: The Learning Analytics Readiness Instrument (LARI)*. Paper presented at the International Conference on Learning Analytics and Knowledge, Indianapolis, IN, USA. doi:10.1145/2567574.2567621
- Colvin, C., Rogers, T., Wade, A., Dawson, S., Gašević, D., Buckingham Shum, S., & Fisher, J. (2015). *Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement*. Retrieved from Sydney, Australia:
- Dawson, S., Joksimovic, S., Poquet, O., & Siemens, G. (2019). *Increasing the Impact of Learning Analytics*. Paper presented at the Proceedings of the 9th International Conference on Learning Analytics & Knowledge, Tempe, AZ, USA. doi:10.1145/3303772.3303784
- Dawson, S., Poquet, O., Colvin, C., Rogers, T., Pardo, A., & Gasevic, D. (2018). *Rethinking learning analytics adoption through complexity leadership theory*. Paper presented at the LAK'18: International Conference on Learning Analytics and Knowledge, Sydney, Australia. doi:10.1145/3170358.3170375
- Ferguson, R., Macfadyen, L., Clow, D., Tynan, B., Alexander, S., & Dawson, S. (2014). Setting Learning Analytics in Context: Overcoming the Barriers to Large - Scale Adoption. *Journal of Learning Analytics*, 1(3), 120-144. doi:<https://doi.org/10.18608/jla.2014.13.7>
- Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology & Society*, 15(3), 42-57.
- Mangaroska, K., & Giannakos, M. N. (2018). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies*. doi:10.1109/TLT.2018.2868673
- Siemens, G., Dawson, S., & Eshleman, K. (2018). Complexity: A Leader's Framework for Understanding and Managing Change in Higher Education. *EDUCAUSE Review*, 53(6), 27-42.
- Tsai, Y.-S., Poquet, O., Gašević, D., Dawson, S., & Pardo, A. (2019). Complexity leadership in learning analytics: Drivers, challenges and opportunities. *British Journal of Educational Technology*, 0(0). doi:10.1111/bjet.12846
- Uhl-Bien, M., Marion, R., & McKelvey, B. (2007). Complexity leadership theory: Shifting leadership from the industrial age to the knowledge era. *The leadership quarterly*, 18(4), 298-318. <https://doi.org/10.1016/j.leaqua.2007.04.002>.

Quantitative ethnography as a framework for network analysis – a discussion of the foundations for network approaches to learning analysis

Morten Misfeldt

University of Copenhagen
misfeldt@ind.ku.dk

Daniel Spikol

University of Malmö
daniel.spikol@mau.se

Jesper Bruun

University of Copenhagen
jbruun@ind.ku.dk

Mohammed Saqr

University of Eastern Finland
mohammed.saqr@uef.fi

Rogers Kaliisa

University of Oslo
rogers.kaliisa@iped.uio.no

Andrew Ruis

University of Wisconsin–Madison
arruis@wisc.edu

Brendan Eagan

University of Wisconsin–Madison
beagan@wisc.edu

ABSTRACT: This workshop explores quantitative ethnography as a framework for discussing network approaches to learning analysis. In many learning contexts, we increasingly have access to large amounts of rich process data. To make meaning of this evidence, our goal is to develop a qualitatively “thick” description of the data, and thus of learning. However, the more data we have, the more difficult this process becomes: qualitative analysis becomes less feasible, and quantitative analysis becomes less reliable. Quantitative ethnography addresses this problem by using statistical techniques to warrant claims about the quality of thick description. The result is a more unified mixed-methods approach that uniquely links the evidence we collect to learning processes and outcomes. This workshop focuses on different techniques that address this challenge, including epistemic network analysis, social network analysis, and Social Learning Analytics. The aim of the workshop is to examine these techniques of network analysis through a quantitative ethnography frame in order to generate a more unified methodology for modeling learning processes and providing actionable insights for research and teaching practices.

Keywords: Quantitative ethnography; network analysis; ENA; SNA; SLA; mixed-methods research

1 BACKGROUND AND PURPOSE

In October 2019, the first International Conference on Quantitative Ethnography (ICQE) brought together scholars seeking to meaningfully analyze and interpret large amounts of rich qualitative data (Eagan, Misfeldt, & Siebert-Evenstone, 2019). Quantitative ethnographic approaches have been used in a variety of fields, including learning analytics, to understand human behavior and interaction.

Quantitative ethnography (Shaffer 2017) views data that document learning processes as evidence about the discourse of particular cultures of learning. To make meaning from this evidence, and thus gain some understanding of learning processes and outcomes, we must strive for what Geertz (1973) called a qualitatively “thick” description of the data. However, the more data that is available, the more difficult this process becomes: qualitative analysis conducted by hand using traditional methods becomes less feasible; at the same time, quantitative analysis becomes problematic because traditional techniques find large numbers of significant results, some with little theoretical grounding and others with very small effect sizes. Quantitative ethnography addresses this problem by using statistical techniques to warrant claims about the quality of thick description. The result is a unified mixed-methods approach that uniquely links the evidence we collect to learning processes and outcomes.

The purpose of this workshop is to continue the discussion begun at ICQE around the use of network analysis, specifically in the field of learning analytics. Using the quantitative ethnography method as framework to organize the discussion, we will explore three different network analytic techniques that integrate qualitative and quantitative discourse analysis (Bruun et al., 2017).

Epistemic network analysis (ENA) models learning processes by constructing networks that represent the cognitive connections learners make in a domain. By modeling patterns of connections in discourse, ENA can help researchers quantify and visualize learning over time for individuals and groups, compare learning across learners or contexts, create trajectories of learning, and model the contributions of individuals to group discourse (Shaffer, Collier, and Ruis, 2016).

Social network analysis (SNA) is concerned with the collection, measurement, and analysis of students’ digital artefacts and online interactions. SNA is used to understand learners’ activities, behaviors, and knowledge creation in a social learning setting (Kaliisa et al., 2019). SNA also been used extensively to investigate large datasets documenting the dynamic, complex, and unique structures of groups in learning contexts. These small groups as a unit can be viewed as complex adaptive systems, in which independent participants interact, self-organize and contribute to a shared understanding of common learning objectives (Saqr et al., 2019).

2 INTENDED OUTCOMES, STRUCTURE, AND ORGANIZATION

The workshop is organized both as a mini-conference and hands-on workshop where the participants (a) present their own learning analytic research grounded in network analytic approaches, (b) learn

about and engage with three network analytic techniques commonly used in learning analytic research in order to compare the different network approaches, and (c) discuss in small groups how the same data could be analyzed with different tools and strategies.

These activities will be grounded in quantitative ethnography and are expected to inform a discussion of the philosophical and methodological foundations for network analysis in learning analytics.

During the workshop:

1. The presenting participants will give a mini-presentation. Abstracts will be pre-circulated, so these presentations are intended to remind participants of the key claims and findings, and provide suggestions to how that abstracts talk to one another.
2. All participants will engage in a hands-on workshop that introduces them to the basic principles and applications of ENA, SNA, and SLA.
3. All participants will be assigned to small groups based on their submissions to explore alternate analytic strategies and discuss the grounding of network approaches to learning analytics in quantitative ethnography.

At the end of the workshop, participants will present the main points of the discussions in the small groups, which will form the basis for a white paper on network analysis in learning analytics.

3 PRESENTATIONS AND DETAILED SCHEDULE

The detailed schedule is shown in the following table:

Time	Activity	Responsible
9:00 - 9:15	Introduction	Morten Misfeldt
9:15 – 9:50	Paper presentations	Andrew Ruis (Chair), Rogers Kaliisa, Luis P. Prieto, Jesper Bruun, Amanda Siebert-Evenstone
10:00 – 10:30	Tool showcase	Brendan Eagan, Kamila Misiejuk, Jesper Bruun, Daniel Spikol, and Mohammed Saqr
10:45 – 11:30	Small group discussions	Rogers Kaliisa and Morten Misfeldt (Facilitators)
11:45 – 12:30	Whole group discussion and next steps	Brendan Eagan and Morten Misfeldt (Facilitators)

The accepted presentations are:

1. “Aligning Quantitative Ethnography and Social Learning Analytics to Understand Online Learning Processes,” Rogers Kaliisa (*Oslo University*)

2. "Epistemic Network Analysis in the Longitudinal Study of Single-Case Lifelong Learning: Benefits for Research and for Learners," Luis P. Prieto, María Jesús Rodríguez-Triana, and Tobias Ley (*Tallinn University*)
3. Similarity Analysis of Nation-Wide Survey on Danish Students' Perceptions of Grades, Jesper Bruun (*University of Copenhagen*)
4. Creating Measurement Spaces: Using Fixed Node Positions to Measure across Data Types, Amanda Siebert-Evenstone (*University of Wisconsin – Madison*)

REFERENCES

- Bruun, Jesper, Mats Lindahl, and Cedric Linder. "Network analysis and qualitative discourse analysis of a classroom group discussion." *International Journal of Research & Method in Education* 42.3 (2019): 317-339.
- Eagan, B., Misfeldt, M., Siebert-Evenstone, A. (2019). *Advances in Quantitative Ethnography: First International Conference, ICQE 2019, Madison, WI, USA, October 20-22, 2019, Proceedings*. Part of the Communications in Computer and Information Science book series (CCIS, volume 1112).
- Geertz, C. (1973). Thick description: Toward an interpretive theory of culture. In *The interpretation of cultures: Selected essays* (pp. 3–30). New York: Basic Books.
- Kaliisa, Rogers, Anders I. Mørch, and Anders Kluge. "Exploring Social Learning Analytics to Support Teaching and Learning Decisions in Online Learning Environments." *European Conference on Technology Enhanced Learning*. Springer, Cham, 2019.
- Saqr, Mohammed, Jalal Nouri, and Ilkka Jormanainen. "A Learning Analytics Study of the Effect of Group Size on Social Dynamics and Performance in Online Collaborative Learning." *European Conference on Technology Enhanced Learning*. Springer, Cham, 2019.
- Shaffer, David Williamson. *Quantitative ethnography*. Cathcart Press, 2017.
- Shaffer, D.W. & Ruis, A.R. (2017). Epistemic network analysis: A worked example of theory-based learning analytics. In C. Lang, G. Siemens, A. Wise, & D. Grasevic (eds.), *Handbook of learning analytics* (pp. 175–187). Society for Learning Analytics Research.

The 2nd Workshop on Predicting Performance Based on the Analysis of Reading Behavior

Brendan Flanagan

Kyoto University, Japan

flanagan.brendanjohn.4n@kyoto-u.ac.jp

Rwitajit Majumdar

Kyoto University, Japan

Atsushi Shimada

Kyushu University, Japan

Hiroaki Ogata

Kyoto University, Japan

ABSTRACT: As the adoption of digital learning materials in modern education systems is increasing, the analysis of reading behavior and their effect on student performance gains attention. The main motivation of this workshop is to foster research into the analysis of students' interaction with digital textbooks, and find new ways in which it can be used to inform and provide meaningful feedback to stakeholders, such as: teachers, students and researchers. Building on the success of last year's workshop at LAK19, this year we will offer participants a chance to take part in a data challenge to predict the performance of 300 students based on the reading behaviors of over 1000 students from the previous year in the same course. Additional information on lecture schedules and syllabus will also enable the analysis of learning context for further insights into the preview, in-class, and review reading strategies that learners employ. Participant contributions will be collected as evidence in a repository provided by the workshop and will be shared with the wider research community to promote the development of research into reading analysis systems.

Keywords: Student Performance Prediction, Data Challenge, Reading Behavior

1 WORKSHOP BACKGROUND

Digital learning materials especially digital textbooks are a core part of modern education, and the adoption of digital textbooks in education is increasing. Digital textbooks and e-books are being introduced into education at the government level in a number of countries in Asia (Ogata et al., 2015). This has prompted research into not only the use of such materials within the classroom, but also the collection and analysis of event data collected from the systems that are used for support and distribution (Flanagan et al., 2017; Ogata et al., 2017; Ogata et al., 2015). In addition to its advantages on students' learning, digital text readers are capable of recording interactions regarding students' reading behaviors. As the materials are read by students using the system, the action events are recorded, such as: flipping to the next or previous page, jumping to different pages, memos, comments, bookmarks, and drawing markers to indicate parts of the learning materials that learners think are important or find difficult.

Despite the increase in use, research analyzing students' interaction with digital textbooks is still limited. Recent review study (Peña-Ayala et al., 2014) revealed that almost half of the papers in Learning Analytics (LA) and Educational Data Mining (EDM) fields are using data from Intelligent Tutoring Systems (ITS) or Learning Management Systems (LMS). Previous research into the reading behavior of students has been used in review patterns, visualizing class preparation, behavior change detection, and investigating the self-regulation of learners (Yin et al., 2015; Ogata et al., 2017; Shimada et al., 2018; Yamada et al., 2017). The analysis of reading behavior can be used to inform the revision of learning materials based on previous use, predict at-risk students that may require intervention from a teacher, and identify learning strategies that are less effective and provide scaffolding to inform and encourage more effective strategies. The digital learning material reader can be used to not only log the actions of students reading reference materials, but also to distribute lecture slides.

The main motivation of this workshop is to foster research into the analysis of students' interaction with digital textbooks, and find new ways in which it can be used to inform and provide meaningful feedback to stakeholders, such as: teachers, students and researchers. This proposal builds upon previous workshops that have focused on student performance prediction based on reading behavior. In previous years at LAK and other international conferences, there have been workshops that have offered open ended data challenges to analyze e-book reading logs and predict the final grade score of learners (Flanagan, 2018; Flanagan, 2019), with 16 and 14 participant submissions respectively. However, to-date the data has been from the same course in one year. This proposal focuses on the realistic scenario of predicting student performance based on data from a previous year in the same course. In the proposed workshop, we will offer a unique opportunity for participants to:

- Analyze large-scale reading log data on over 1,000 students with performance-based labels for model training from one course.
- Examine anonymized reading log datasets from a course from the previous year to predict student performance of around 300 students in the same course in the following year.
- Investigate preview, in-class, and post-class reading behaviors by analyzing the scores from quizzes/exams/final grades, lecture schedules and syllabus information that will be provided as part of the datasets.

This year we will provide two datasets: a labeled training dataset of over 1000 students and a test dataset with around 300 students data from the next year's classes. The learner's performance score for the test dataset will be withheld, and participants can upload their scores to the workshop website to check the results of the evaluation once per day. A leaderboard will be provided with the

best evaluation score that each participant has achieved to encourage competition between teams. Compared to the previous year's class, only small updates have been made to the reading materials, offering a real-world scenario for participants to tackle the problem of performance prediction based on digital reader usage.

2 OBJECTIVES

An emphasis was placed on the following topic as the main objective of the workshop, which the organizers feel attention should be paid. Low retention and high failure rates are important problems in education (Villagr -Arnedo et al., 2017). However, studies have shown that timely interventions for at-risk students can be effective in helping change their behaviors (Arnold et al., 2012; Tanes et al., 2011). Therefore, focusing on the early detection of at-risk students is an essential step to changing student's behavior for greater success.

This broader task may be approached from the following perspectives:

- Student reading behavior self-regulation profiles spanning the entire course
- Preview, in-class, and review reading patterns
- Student engagement analysis; and behavior change detection
- Visualization methods to inform and provide meaningful feedback to stakeholders

Participant contributions, such as: paper, programs, source code, will be collected as evidence in a repository provided by the workshop and will be shared with the wider research community to promote the development of research into reading analysis systems. Also, there is an opportunity to integrate the results as part of an ongoing open learning analytics tool development project for inclusion as an analysis feature. This integration of research conducted in the proposed workshop into open learning analytics infrastructure will be managed by the organizers as an ongoing effort.

The contributions to data challenge workshops can potentially be integrated into learning analytics platforms in a number of different capacities. Analysis and the creation of prediction models for at risk students could be integrated as: a passive widget on a dashboard with the option to intervene manually by sending a message to the student in question, or as a fully automated or teacher moderated process where interventions are actively undertaken by the system whenever a at risk student is predicted. Some more recent studies into at risk students not only focus on the prediction, but also on possible causes and interpretation of the conditions that lead to the student being at risk of failure. These works could be intergraded to provide meaningful intervention and help both teachers and students interpret the situation and seek out plausible actions that could be taken to overcome difficulties. The visualization of characteristic reading behavior that lead to high learner performance is also an important topic, such as the research by Minematsu et al. (2019) that investigated the behaviors of power users.

ACKNOWLEDGEMENTS

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (S)16H06304 and NEDO Special Innovation Program on AI and Big Data 18102059-0.

REFERENCES

- Arnold, K. E., & Pistilli, M. D. (2012, April). Course signals at Purdue: Using learning analytics to increase student success. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 267-270). ACM.
- Flanagan, B., & Ogata, H. (2017). Integration of Learning Analytics Research and Production Systems While Protecting Privacy. In *International Conference on Computers in Education (ICCE2017)* (pp. 333-338).
- Flanagan, B., Chen, W., & Ogata, H. (2018). Joint Activity on Learner Performance Prediction using the BookRoll Dataset. In *International Conference on Computers in Education (ICCE2018)* (pp. 480-485).
- Flanagan, B., Shimada, A., Yang, S., Chen, B.L., Shih, Y.C., & Ogata, H. (2019). Predicting Performance Based on the Analysis of Reading Behavior: A Data Challenge, *Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge*.
- Majumdar R., I Boticki., Ogata H. (2019). International Workshop on Technology-Enhanced and Evidence-Based Education and Learning, *Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge* (pp. 595-599).
- Minematsu, T., Shimada, A., Taniguchi, R. (2019). Analytics of the relationship between quiz scores and reading behaviors in face-to-face courses. In *Companion Proceedings 9th International Conference on Learning Analytics & Knowledge (LAK19)*.
- Ogata, H., Taniguchi, Y., Suehiro, D., Shimada, A., Oi, M., Okubo, F., Yamada, M., & Kojima, K. (2017). M2B System: A Digital Learning Platform for Traditional Classrooms in University. *Practitioner Track Proceedings* (pp.155-162).
- Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-Book-based learning analytics in university education. In *International Conference on Computer in Education (ICCE 2015)* (pp. 401-406).
- Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, 41(4), 1432-1462.
- Shimada, A., Taniguchi, Y., Okubo, F., Konomi, S., & Ogata, H. (2018). Online change detection for monitoring individual student behavior via clickstream data on E-book system. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18)*. (pp. 446-450).
- Tanes, Z., Arnold, K. E., King, A. S., & Remnet, M. A. (2011). Using Signals for appropriate feedback: Perceptions and practices. *Computers & Education*, 57(4), 2414-2422.
- Villagr -Arnedo, C. J., Gallego-Dur n, F. J., Llorens-Largo, F., Comp n-Rosique, P., Satorre-Cuerda, R., & Molina-Carmona, R. (2017). Improving the expressiveness of black-box models for predicting student performance. *Computers in Human Behavior*, 72, 621-631.

Improving Learning Analytics and Student Performance through Connected Lifelong Learning on the Blockchain

Ocheja Patrick, Flanagan Brendan, Lecailliez Louis and Ogata Hiroaki

Kyoto University

ocheja.ileanwa.65s@st.kyoto-u.ac.jp; flanagan.brendanjohn.4n@kyoto-u.ac.jp;
louis.lecailliez@outlook.fr; ogata.hiroaki.3e@kyoto-u.ac.jp

ABSTRACT: The use of learning analytics to enhance learning at different levels has continued to gain attention over the past few years. With learning activities taking place in different environments, systems and contexts, capturing and sharing these actions/outcomes continue to pose serious problems. In this paper, we examine how student learning interactions and events across various learning systems such as BookRoll can be accessed, transferred and protected across different learning environment. In facilitating the reusability of past learning interactions, we propose the transferability of user models derived from these learning logs. Finally, we discuss potential areas for advancing the field in connecting and using distributed learning logs for improving learning analytics and students' performance in general.

Keywords: lifelong learning, learning analytics, user model, privacy, learning record store, learning management system, blockchain, smart contract

1 INTRODUCTION

It is common to experience learning in different ways ranging from informal to formal contexts, and passive to active engagements. However, due to the ubiquitous nature of learning, it is difficult to capture and unify learner data from different environments as commonly experienced in other Big Data environments (Kadadi, Agrawal, Nyamful & Atiq, 2014). In this paper, we focus on discussing challenges to connecting lifelong learning data of students across different learning environments. Lifelong learning log is a journal that contains all of the learning activities carried out by a learner, and consists of a sparse multisource dataset of the learning actions of a learner. As learners change school, it is important to enable learning traceability by connecting learning experiences, and revisit consistent stakeholder concerns on data privacy and security. We demonstrate how lifelong learning can be achieved using blockchain technology and present some results from a live deployment in a K-12 environment.

1.1 Related Work

Lifelong learning is desirable and useful for learning analytics (Mouri & Ogata, 2015; Bakharia, Kitto, Pardo, Gašević & Dawson, 2016) as it provides ways to understand what a learner knows beyond current assessment and probe in detail the origin of difficulties or excellence. Due to the limited solutions that can facilitate data continuity across different learning environments, the act of combining data from multiple sources becomes a common alternative when lifelong learning data of learners is difficult to obtain or store.

Samuelson, Chen, & Wasson (2019) in a review on multisource data for learning analytics identified the lack of tools or research work on meaningful data integration, storage and processing when combining multisource data. Kay & Kummerfeld (2019) proposed a conceptual model for evaluating how learning applications and data repositories can be used to realize Personal User Model for Lifelong, Life-wide Learners (PUMs). PUMs is proposed as a personal repository of raw data and interpretations that could be accessed by authorized programs. One question that comes up when implementing PUMs is how to integrate with existing tools. Kay & Kummerfeld (2019) acknowledged this concern and mentioned interoperability as one of the requirements for realizing PUMs. Another challenge with the PUMs framework is how to manage privacy especially when user models are centrally stored and learners or their institutions may become less aware of how such models are been used.

To solve the limitations of centralizing lifelong learning, Ocheja, Flanagan & Ogata (2018) and Ocheja, Flanagan, Ueda & Ogata (2019) proposed a decentralized architecture where Learning Record Stores (LRS's) containing learner data are connected together through the blockchain and privacy permissions are managed using smart contracts. From the outcome of previous implementations, this paper discusses potential solutions to overcome the challenge of sharing and reusing learning experiences for analytics.

2 SOLUTIONS TO CURRENT PROBLEMS

As identified in the review of existing works on facilitating and enabling lifelong learning in the previous section, there is consistent concern with how to manage resulting data from learning tools, maintain privacy of learner information and facilitate learning analytics. In this work, we propose possible solutions to the issues of connecting distributed learning infrastructure, managing learning data, transferring user models and ensuring privacy.

2.1 Decentralization

As learning is ubiquitous in nature, and therefore realizing lifelong learning should not be the sole responsibility of one organization, but rather a function of places in which learning occurs. To capture activities of learners in different environments and institutes, Ocheja, Flanagan & Ogata (2018) proposed a decentralized learning analytics platform. A decentralized network makes it possible for multiple systems controlled by different actors to interact and transparently reach a consensus on protected resources. For example, to offer more privacy control on PUMs (Kay & Kummerfeld, 2019), it becomes necessary to decentralize the user model such that when students change institution, they move with their model and update it at their new school or learning environment as new learning activities occur. For e-book learning logs, a decentralized access to lifelong learning makes it possible to obtain additional information outside the e-book context but impactful on learning outcome.

2.2 Tracing Learning

As learners move from one institute to another, it becomes necessary to know which institutes they have been to previously. One reason for such requirement is a case where a teacher needs to trace the root cause of a particular difficulty experienced by their student. In figure 1, *Bob's* teacher is faced with the task of detecting the gap in *Bob's* past learning in a prerequisite course. To detect this gap,

Bob's teacher needs to know what topics in *Statistics* were covered at Bob's previous school and what Bob's performance and mastery was in each of these topics, which could be found in the learning logs from previous education institution. Connecting learning records on the blockchain provides an additional benefit of enabling traceability. This can be achieved using the nested transactions feature which is fundamental function of the design in blockchain where the current block contains a reference to the previous block.

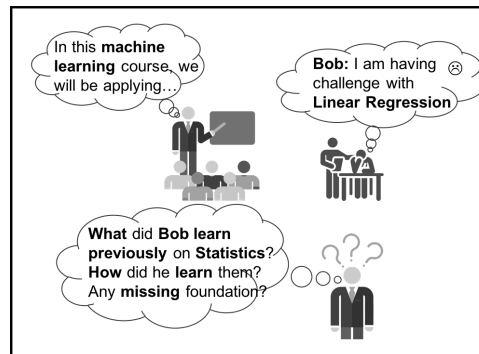


Figure 1: Tracing a learner's learning path.

2.3 Privacy

The lack of protection and control of private information by data owners exist as a result of the disconnection between different LRS's. An example of this problem can be seen when students move from one school to another and in the process are less aware of how their past learning data is being used. Although, learning analytics helps in improving the performance of learners (Okubo, Yamashita, Shimada & Ogata, 2017), the gains of learning analytics must be commensurate to respecting learner's privacy and associated rights (Rubel & Jones, 2016). While connecting learning logs across different systems and engendering transfer of these logs, it is necessary to prioritize learner's privacy: learners should be constantly aware and have control of their data. As proposed system connects all of a learner's data across institutes, it allows learners to still control their data at previous institutions through use requests even after they have ended their formal association with an education institution. This also enables the possible reuse of data for research purposes by providing a method of formally requesting access and use of a learner's data even after they have left an institution.

2.4 Shareable and Reusable Learner Models

Connecting and transferring learning data can be useful for creating learner models, however under-resourced institutions are often at a disadvantage because of data and computational limitations. By enabling sharing of models and data between institutions it may support the realization of more accurate learner models, and therefore it should be possible to allow learners to transfer these models across different institutes where they are enrolled. This will help to advance the learning personalization process and reduce the computational cost and technical requirements for recomputing a learner's model whenever they change school (Baker, 2019).

3 BOLL AS A TOOL FOR LIFELONG LEARNING AND ANALYTICS

We present the Blockchain of Learning Logs (BOLL) proposed in Ocheja et al. (2018 & 2019) and demonstrate how it enables lifelong learning and analytics with results from a live deployment at K-12 school. BOLL is a platform that enable learners to connect their lifelong learning events as verifiable and non-modifiable transactions on the blockchain. In a practical sense, learners can move from one learning institution to another and at the same time take all their past learning actions with them. We consider this solution a notable example as it provides answers to questions on privacy, decentralized access, transfer of records, tracing learning, and openness for integration. As BOLL is currently in active development, we present to the research community a framework that serves as a recipe for enabling lifelong learning and at the same time open for collaboration.

BOLL in one stretch, solve problems on data privacy, trust between stakeholders and third parties and the overhead of manually transferring and collecting learner data across different institutions and systems. The use of a decentralized technology effectively ensures that no single party can abuse the interest of others without having at least the consent of 51% or more members on the network (Nakamoto, 2008). In this light, the questions we now ask revolve around how to on-board various institutions, facilitate a seamless transition and advance the field along this path.

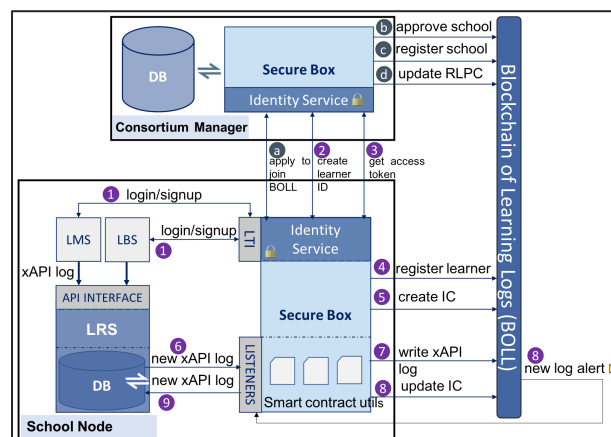


Figure 2: BOLL enrolment procedure.

3.1 Onboarding Requirements

Learning organizations can join BOLL and become stewards to safeguarding the learning records of learners across all institutes on the network. These learning organizations are regulated through a consortium which ensure members are verified and accountable. In addition to existing learning technologies, each institute is required to have an Ethereum blockchain node and a SecureBox: an open source by application project containing a set of utilities for interacting with the blockchain.

As for learners, they can join BOLL through their host institution. Because each learning institute has its own authentication system, BOLL allows each institute to connect to their authentication system using a Learning Tools Interoperability (LTI) module. This procedure is mandatory at least for the first-time a learner accesses BOLL and processes 1–5 in figure 2 are carried out. Subsequent access to BOLL is authorized through OAuth2 provided by the consortium manager. In a case where learners are underage and require parental consent (e.g. K-12 learners), BOLL provide settings for automatically

generating the accounts for such learners when their parent or guardian opts-in to logging their ward's learning actions on BOLL. In figure 3, we show a distribution of some on-boarding processes from a live deployment of BOLL at a K-12. Creating smart contract for records indexing, registering user and pairing learners to institutes are done automatically for the K-12 learners.

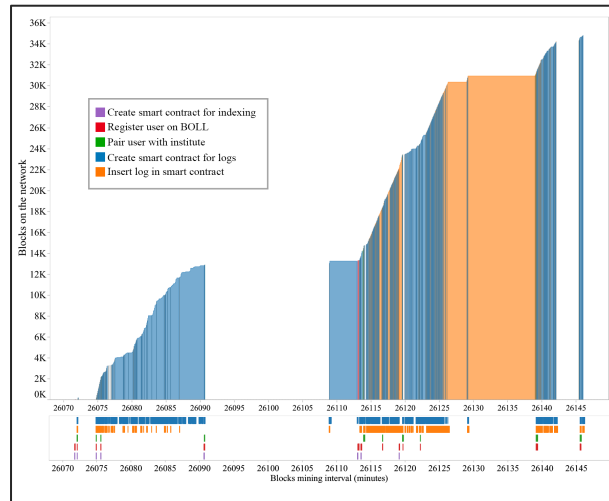


Figure 3: Snapshot of BOLL network deployed at K-12.

3.2 Privacy: Access and Authorization

BOLL facilitates learner privacy by using smart contracts as proposed by Ocheja et al. (2018) and implemented by Ocheja et al (2019). BOLL groups learning data according to the action verb that denotes the learning action performed. BOLL then assigns a specific smart contract to each type of learning action. A learner may grant a *read*, *write* and/or *admin* permissions to another party. A read permission allows the party to view the learner's records. A write permission allows the party to write learning logs on behalf of the learner. Only approved learning institutes can write these data for learners. An admin permission gives the party access to read and write learning actions for the learner and also allows the party to give other parties similar permissions on the learner's data. In the current implementation of BOLL, permissions are based on action verbs of the IMS Caliper (IMS Caliper, 2013) and/or the Experience API (xAPI) (Advanced Distributed Learning, 2016) specifications. BOLL recommends the use of open learning logs standards such as the xAPI and IMS Caliper.

3.3 Creating and Viewing Learning Logs

The task of creating learning logs on BOLL is done in the background while learners interact with learning tools connected to BOLL. As shown in steps 6 and 9 on figure 2, listeners are configured to pick up new learning events and write them to the blockchain or another institute's LRS. To issue testimonials such as certificates, recommendation letters and other documents, BOLL provides a view for staff members of an institution to create and issue such documents. All issued testimonials contain the cryptographic signature of the issuer. Most of the transactions from the live deployment of BOLL result from creating and inserting learning events as shown in figure 3. BOLL also provides an interface for learners to view their learning logs at different institutions.

3.4 Connect and Transfer Learning Logs

Connection of learning data is enabled on BOLL by writing each learning action on the blockchain. When a record is written on the blockchain for a learner, all institutes that have permission to read their learning records are notified through global events emitted on the blockchain. These institutes can then request for a copy of the new learning log from the originating institute's LRS. To transfer past learning records to a newly approved institute, the SecureBox contain functions that can query past LRS's where a learner has previously schooled to get their past data. These functions are automatically triggered by a *permission-granted* event immediately the learner grants access to the new institute.

3.5 Making Sense of Lifelong Learning

BOLL provide some useful features for understanding a learner's past learning experience. For example, a teacher can ask their student for permission to view the student's lifelong learning on BOLL. When the permission is granted, the teacher can view the courses their student has previously taken. In order to understand how their past learning could be related or useful to the teacher's course, the teacher can pick one of several models to run on the student's past logs. This view is shown in figure 4. The teacher first selects which course on the LMS or learning tool they want to view student's lifelong learning. After selecting the student, the teacher can then select which of the learner's past schools to get the learner's lifelong learning. Because the past learning events can be of different types, the teacher can specify which type of learning actions to include in the retrieved data. The final step is to decide which learning analytics model to apply on the retrieved data. Here, we identify a potential for BOLL to integrate with other learning analytics dashboard and visualization tools such as that proposed by Majumdar, Akçapınar, Akçapınar, Ogata & Flanagan (2019).

The screenshot shows a web interface for analyzing student data. At the top, there are two tabs: 'My Logs' and 'My Students' Logs'. Below the tabs is a section titled 'Analyze student's data'. This section contains four input fields arranged in a 2x2 grid. The top-left field is 'Course' with a dropdown menu labeled 'Select Course'. The top-right field is 'Student' with a search input labeled 'Start typing student's name to search' and a magnifying glass icon. The bottom-left field is 'School' with a dropdown menu labeled 'Select School'. The bottom-right field is 'Type of Learning Log' with a search input labeled 'Start typing to search...' and a magnifying glass icon. Below these fields is an 'LA Model' dropdown menu labeled 'Select model to run on data'. To the right of the 'LA Model' dropdown is a blue button labeled 'Load Data'.

Figure 4: BOLL interface for Analyzing Lifelong Learning.

Another example of how BOLL makes sense of lifelong learning is in helping teachers understand a learner's knowledge map. For example, a teacher may want to know what a learner knows in relation to a particular subject. A learner's knowledge map can then be constructed with their lifelong learning on BOLL. Thus, BOLL can serve as a data backend for knowledge map analysis tools like the tool proposed by Flanagan, Majumdar, Akçapınar, Wang & Ogata (2019). Flanagan et. al (2019) proposed a knowledge map creation platform capable of computing a learner's knowledge states over time, grouping knowledge states, intra-cohort comparison and computing relationship between various knowledge states. Such visualizations provide a quick and concise view of what a learner knows and what they may find easy or difficult to grasp.

3.6 Metadata: Public and Private

When transactions representing learning actions of students are written on the blockchain, all participants on the network get notified. But because the BOLL network does not store the actual learning actions on the blockchain, it is not possible to see what learning actions learners performed. However, when some metadata such as test scores or grades are stored directly on the blockchain, the decision on whether to make such contents public or private becomes a concern for stakeholders. One way to ensure that scores which are expected to be kept private remain so is that such scores should not be included as metadata. Another alternative could be to redact such scores by using some non-publicly communicated offsets or encryption before storing them on the blockchain. For a consortium, it is more appropriate to provide standards to guide all members but some institutions may decide on what works best for them. In this case, any choice made should be communicated to other parties that want to make sense of such data.

4 FUTURE IMPLICATIONS FOR THE FIELD

As the field of lifelong learning and analytics continues to evolve, the following important topics remain and require investigation in greater detail. By extending the framework proposed in this paper, it may be possible to overcome part of these problems, and it would require more collaboration with stakeholders and the wider research community.

Transferring and connecting learner/learning models: When a learner's learning data is used to construct a model, it is useful to enable the learner to keep a record of such model and be able to grant their subsequent learning institutions access to such models. One of the arguments in favour of this feature is a case where replicating the same model might be unachievable especially when different institutions have varying access to different learning analytics technologies. Thus, it is necessary to understand the requirements for sharing learner/learning models and enable such possibilities when connecting lifelong learning.

Off-the-shelf learning analytics for connected lifelong learning: Connecting lifelong learning becomes useful if teachers and/or students can obtain meaningful feedback from the connected data. To help them gain insights from these connected data, it is necessary to provide tools for analytics and visualization. While there are existing efforts in providing learning analytics dashboards, future concerns should be focused on decentralizing learning analytics dashboards such that they can be easily integrated with data coming from multiple institutions. Also, because the connected lifelong learning data could be of different types (xAPI standard, IMS Caliper standard, etc.) it is also necessary to develop a framework to unify these different standards.

Integrating different learning logs standards: As lifelong learning involves learning actions of a learner at different schools where different learning technologies are used, it is necessary to ensure that such learning actions are compatible or can be merged. A common way of ensuring compatibility is the adoption of a standard such as xAPI for expression learning actions. However, it is possible that some institutions may prefer a set of learning events standards over another. Therefore, we recommend the development of a framework to unify learning logs resulting from different standards. While this task might be trivial in a case where learning actions are expressed in well-defined standards, this task becomes more difficult when no proper documentation on the format of learning actions are

provided. A community driven approach could be to maintain a public repository that list all available standards for learning actions stored on BOLL. This makes it possible for learning analytics tools to conveniently interpret connected lifelong learning regardless of the variation in data format.

Demise of an institution: As only a hash of the learning log and its location is recorded on BOLL, there is a possibility of a learning log outliving its host institution. To illustrate this, a student might graduate from an institution and 10 years later, that institution ceases to exist. In a case where all computing facilities such as the LRS of that institution is shutdown, then the learning logs whose references are held on the blockchain cannot be retrieved anymore. To solve this problem, it will be helpful to have a learning blockchain with third parties who can offer data backup services.

5 CONCLUSION AND FUTURE WORK

In this work, we assessed various contributions in the field of learning analytics towards the realization of connected lifelong learning and use of multisource data to improve learning. A review of previous research shows that being able to access lifelong learning logs is useful for learning analytics. But this is still lacking in the field especially due to multiplicity of learning tools, security and privacy concerns, and the administrative burden of using and managing existing solutions.

To solve the problem of disconnected lifelong learning and challenges with adoption of potential solutions to existing problems, this paper presented new pivots that can be used to advance the field of learning analytics towards the realization of connected lifelong learning. One of such pivots is the decentralization of learning analytics and tools such that learning organizations can operate as smaller units in a large network with lesser administrative burden when deciding rules on data collection, access, usage and transfer across institutes. The BOLL network is presented as a solution to an existing challenge of ensuring trust among institutions and the difficulty in deploying learning analytics on a wider scope when policies are continuously evolving. Future work will be focused on providing answers to challenges presented as implications for the field including transfer of learning models, integrating off-the-shelf learning analytics, unifying learning logs standards and conducting user studies to measure the impact of the proposed BOLL system.

ACKNOWLEDGEMENT

This research was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (S) Grant Number 16H06304.

REFERENCES

- Advanced Distributed Learning. 2016. Experience API (xAPI) Specification. <http://github.com/adlnet/xAPI-Spec>. Accessed September 10, 2019.
- Baker, R. S. (2019). Challenges for the Future of Educational Data Mining: The Baker Learning Analytics Prizes. *JEDM/ Journal of Educational Data Mining*, 11(1), 1-17.
- Bakharia, A., Kitto, K., Pardo, A., Gašević, D., & Dawson, S. (2016). Recipe for success: lessons learnt from using xAPI within the connected learning analytics toolkit. *In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 378-382). ACM.

- Flanagan, B., Majumdar, R., Akçapınar, G., Wang, J., & Ogata, H. (2019). Knowledge Map Creation for Modeling Learning Behaviors in Digital Learning Environments. *In Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge (DC@ LAK19)*.
- IMS Global Learning Consortium. (2013). Learning measurement for analytics whitepaper. *IMS Learning Analytics White paper*, 1-12.
- Kadadi, A., Agrawal, R., Nyamful, C., & Atiq, R. (2014). Challenges of data integration and interoperability in big data. *In 2014 IEEE international conference on big data* (pp. 38-40). IEEE.
- Kay, J., & Kummerfeld, B. (2019). From data to personal user models for life - long, life - wide learners. *British Journal of Educational Technology*, 50(6), 2871-2884.
- Majumdar, R., Akçapınar, A., Akçapınar, G., Ogata, H., & Flanagan, B. (2019). LAViEW: Learning Analytics Dashboard Towards Evidence-based Education.
- Mouri, K., & Ogata, H. (2015). Ubiquitous learning analytics in the real-world language learning. *Smart Learning Environments*, 2(1), 15.
- Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.
- Ocheja, P., Flanagan, B., & Ogata, H. (2018). Connecting decentralized learning records: a blockchain based learning analytics platform. *In Proceedings of the Eighth International Conference on Learning Analytics & Knowledge* (pp. 265-269). ACM.
- Ocheja, P., Flanagan, B., Ueda, H., & Ogata, H. (2019). Managing Lifelong Learning Records Through Blockchain. *Research and Practice in Technology Enhanced Learning*, 14(1), 4.
- Okubo, F., Yamashita, T., Shimada, A., & Ogata, H. (2017). A neural network approach for students' performance prediction. *In International Conference on Learning Analytics and Knowledge* (pp. 598-599).
- Rubel, A., & Jones, K. M. (2016). Student privacy in learning analytics: An information ethics perspective. *The Information Society*, 32(2), 143-159.
- Samuelson, J., Chen, W., & Wasson, B. (2019). Integrating multiple data sources for learning analytics review of literature. *Research and Practice in Technology Enhanced Learning*, 14(1), 11.

Design of an Interactive Dashboard for Teacher Orchestration in Collaborative Science Inquiry

Jiaxin Cao and Yanjie Song
The Education University of Hong Kong
jxcao@s.edu.hk

ABSTRACT: This paper presents the design of a theory-led and interactive dashboard for teacher orchestration in collaborative science inquiry for primary school education. By aligning the design with the collaborative science inquiry model and theoretical principles of class orchestration, the dashboard visualizes collaborative inquiry behaviors of students and teacher's feedback in a multi-track view holistically and simultaneously. Instructors can follow the grasped information to orchestrate collaborative learning. The dashboard visualization aims at supporting instructors' pedagogical decision-making to orchestrate collaborative learning by multiple types of feedback in different inquiry phases at varied social levels.

Keywords: Learning Analytics, Interactive Dashboard, Teacher Orchestration, Collaborative Science Inquiry

1 INTRODUCTION

Collaborative inquiry learning is one of the most challenging and exciting ventures for today's schools (Phua, 2013). Teachers need support for accomplishing orchestrating learning among complex factors. The learning dashboard has the potential to address the difficulty by supporting the judgment, inference, and decision-making of teachers (Alhadad, 2018). However, the majority of learning dashboards still uses basic charts, graphs, or scatterplots without actionable information for guiding teaching (Schwendimann et al., 2017). This paper proposes to adopt theory-led principles to design an interactive dashboard that takes into account the pedagogical model to make learners' collaborative inquiry process visible and supports teachers to orchestrate collaborative science inquiry through an interactive and simultaneous flow of data for real-time visualization.

2 LITERATURE REVIEW

2.1 Teaching in Collaborative Science Inquiry

Collaborative science inquiry refers to "the demand of practicing inquiry in science education and the increasing proliferation of computer-supported collaborative learning (CSCL) (Bell, Urhahne, Schanze, & Ploetzner, 2010, p. 349)". In inquiry-based learning in science, students act as "scientists" to identify problems, make plans, and solve problems (Bransford, Brown, & Cocking, 2000) rather than receive knowledge from the teacher passively in the traditional way (Crawford, Kelly, & Brown, 2000). On the other hand, the technical evolution creates more opportunities to work in small groups and to engage in scientific inquiry (Oliveira, Boz, Broadwell, & Sadler, 2014), such as to foster their motivation in collaboration (Luis et al., 2011). Science educators are increasingly embracing collaborative inquiry learning because it focuses on developing scientific thinking and collaborative

skills, such as questioning, critical thinking, and problem-solving (Savery, 2015), which are aligned with the competencies in the 21st century (Boyatzis & Boyatzis, 2008).

In collaborative science inquiry, students work more on their initiatives and are less dependent on teachers (Bell et al., 2010). Teachers act as facilitators. However, to guarantee learning occurs in a small-group inquiry is not an easy task in reality (Webb, 2009), as the students may be off their learning task without teacher guidance in time. Thus, the role of a teacher is essential in the success of collaborative inquiry-based learning in science, and the teacher is expected to redirect the group work in a productive direction by the pedagogical intervention (Dillenbourg, 1999).

Some instructional models summarized these interventions in students' inquiry and collaboration, such as (1) 5E model with Engage, Explore, Explain, Elaborate, and Evaluate (Bybee & Landes, 1990; Duran & Duran, 2004), (2) Engage, Explore, Explain, and Extend (Marshall & Horton, 2011). However, teachers might still fail to shape pedagogical decisions in their teaching practice because the progression of inquiry activities, in reality, is not necessarily linear and typical (Kwon et al., 2018). If there is not a method to guide teachers, they may feel frustrated in facilitating students' collaborative science inquiry among different social levels (e.g., class, group, and individual) synchronously (van Leeuwen, 2015).

2.2 Teacher Orchestration: In and Beyond Classroom

The metaphor of class orchestration could be an effective strategy to help teachers cope with difficulties in design and real-time management of multiple learning activities (Dillenbourg & Jermann, 2010). A common English dictionary defines "orchestrate" as "to arrange or combine so as to achieve a desired or maximum effect" (Merriam Webster Inc., 2019). Class orchestration (1) conveys technical and pedagogical flavors, (2) describes a teacher as a driver of classroom activities with constructivist approaches supported by ICT (Information Communication Technology), instead of a conductor with more teacher-centered lecturing, and (3) advocates teacher management of real-time and multi-layered activities in a multi-constraints context (Dillenbourg, 2013). Thus, class orchestration can be considered as a process that enables the teacher to productively coordinate supportive interventions (Fischer & Dillenbourg, 2006).

The teacher orchestrating learning in science education represents a spectrum of congruent approaches (Watts, 2003). Dillenbourg and Jermann (2010) also pointed out the affordances of orchestration to indicate the pedagogical strategies that work well, which include flexibility, continuity, awareness, curriculum relevance, assessment relevance, and collaboration. However, the previous models of orchestration, namely class orchestration, have focused more on in-class learning activities. In the context of CSCL, the development of technologies (e.g., mobile and social media) leads to more complex and dynamic learning environments. Collaborative science inquiry, powered by CSCL, has already exceeded the walls of a classroom. Hämäläinen, Kiili, and Smith (2017) found that it would be problematic to remain orchestration in the classroom but not across different learning settings. A technical solution guided by a collaborative inquiry-based learning approach should integrate these orchestration affordances to meet the needs of the primary role – teacher, but not limited in the context – classroom, to bring about conducting collaborative science inquiry in an orchestrated way.

2.3 Theory-led and Interactive Dashboard

The learning dashboard is a typical pipeline of learning analytics to transform educational data to inform pedagogical decisions (Podgorelec & Kuhar, 2011). There is a critical difference in the use of data between learning analytics and traditional educational research (Dragan, Shane, & Abelardo, 2016). However, a dashboard always provides abundant data in real-time and might overwhelm users (Charleer, Klerkx, & Duval, 2014). Lockyer, Heathcote, and Dawson (2013) point out that effective data presentation requires a thorough understanding of the pedagogical and technical context. Theory-oriented visualization strategies align educational theories by organizing visual elements to address educational contexts holistically (Cao & Song, 2019) and make the understanding of concerns explicit and upfront (Kelly, Thompson, & Yeoman, 2015). Interactive visualization enables users' interaction with the model and data (Sacha et al., 2017) according to their cognitive and decision-making needs. Therefore, the design of an interactive dashboard led by the theory-led pedagogical model obtains the mechanism to provide the affordances of teacher orchestration and make the collaborative science inquiry process visible for flexible pedagogical actions.

3 DESIGN OF INTERACTIVE DASHBOARD

3.1 Need for an Interactive Dashboard

Actually, teachers could encounter difficulties during orchestrating collaborative science inquiry. The collaborative inquiry learning behaviors are distributed in various forms and places. These behaviors are usually created by multiple stakeholders on specific phases or social levels so that it is hard to obtain a bird's-eye view of them in a glance. After knowing, teacher still need further supports in, namely, (1) recognizing problems and needs of each group and even whole class just in time, (2) identifying appropriate strategies to address the issues and needs, (3) reviewing the impact of strategy on students' inquiry behavior, and (4) making sense of what to facilitate in the next step. Then, teachers may be able to provide feedback (such as tips and comments) for students according to the information.

3.2 Briefing of the Purposed Interactive Dashboard Design

The purposed dashboard (in *Figure 1*) is guided by the collaborative science inquiry model and orchestration principles to address the teachers' demands mentioned above. All inquiry behaviors in the five inquiry phases are presented in different tracks were overlapping and non-linear processes visualized. Dragging the sliding bar on the top of right can zoom in or out the timeline. In each track, the upper part presents both the circle icon as macro-feedback (in classroom scope) and the tag icon as micro-feedback (in specific group scope). The feedback is in three types – resource, comment, and evaluation. In its lower part, horizontal bars nest a series of inquiry behaviors at the same time. The bars could be clicked to view the specific activities in a pop-up bubble. Then, teachers could interact with the dashboard by (1) dragging and arranging feedbacks among five phases with continuity, (2) checking inquiry behaviors to raise awareness of student's activity state, (3) deploying three types of feedbacks both in advance and on the fly with flexibility, and (4) releasing feedbacks with integration of scenario in both micro- and macro-feedback.

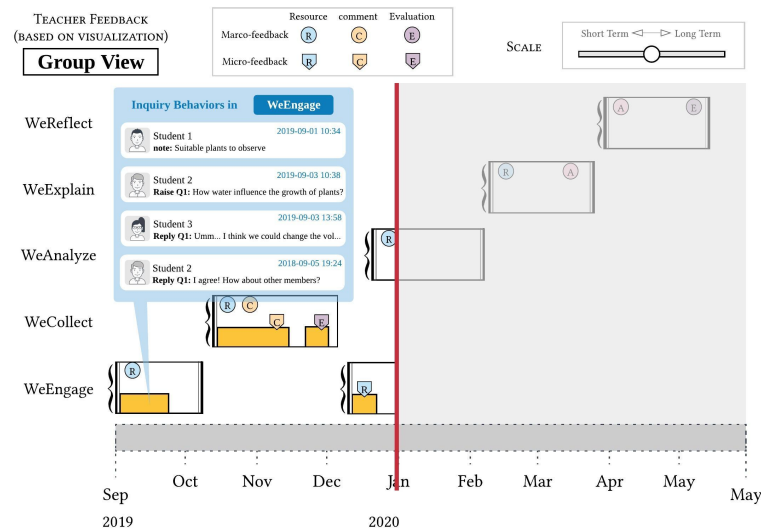


Figure 1: Demonstration of group view in the interactive dashboard

3.3 Deployment of the Interactive Dashboard in m-Orchestrate Learning System

The m-Orchestrate learning system aims at manipulating teachers orchestrating the collaborative science inquiry (Song, Cao, Tam, & Looi, 2019), which supports real-time interactions among teachers and students premised on a collaborative inquiry model. The model contains five phases, namely, WeEngage, WeCollect, WeAnalyze, WeExplain, and WeReflect, and each equipped with specific plugins for supporting inquiry needs (see *Table 1*). The system obtains a clear scheme to make data in different inquiry states relocatable in the holistic process of inquiry. Thus, it is a suitable platform to be the data source and deploy the proposed dashboard.

Table 1: Collaborative inquiry behaviors in the m-Orchestrate learning system.

Inquiry Phase	Behavior (Bell et al., 2010)	Reflecting on the m-Orchestrate learning system
WeEngage	searching for information	organizing/sharing background information by note/comment
	asking question	asking some possible questions for inquiry
	hypothesis generation	raising inquiry questions
WeCollect	planning tasks	planning inquiry tasks in the to-do-list plugin
	work division	conducting inquiry tasks in the to-do-list plugin
	investigation	recording data by note/spreadsheet plugin
WeAnalyze	analyzing data	calculating/processing data by note/spreadsheet plugin
	interpreting data	interpretation by note/chart drawing plugin
	modeling	outlining models by mind map plugin (included in the note)
WeExplain	conclusion	drawing conclusion by notes
	evaluation	finishing the tests released by quiz system
WeReflect	reflection	reflecting the whole inquiry process by notes and comments
	prediction	predicting and inferring other phenomena by note/comment

4 FUTURE WORK

In the future, we are going to develop the interactive dashboard and then conduct investigations to evaluate the effectiveness of the interactive dashboard in supporting teacher orchestration practice in collaborative science inquiry.

REFERENCES

- Alhadad, S. S. (2018). Visualizing Data to Support Judgement, Inference, and Decision Making in Learning Analytics: Insights from Cognitive Psychology and Visualization Science. *Journal of Learning Analytics*, 5(2), 60-85.
- Bell, T., Urhahne, D., Schanze, S., & Ploetzner, R. (2010). Collaborative inquiry learning: Models, tools, and challenges. *International journal of science education*, 32(3), 349-377.
- Boyatzis, R., & Boyatzis, R. E. (2008). Competencies in the 21st century. *Journal of management development*.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn* (Vol. 11): Washington, DC: National academy press.
- Bybee, R. W., & Landes, N. M. (1990). Science for Life & Living: An Elementary School Science Program from Biological Sciences Curriculum Study. *American Biology Teacher*, 52(2), 92-98.
- Cao, J., & Song, Y. (2019). *Beyond Just Following Data: How Does Visualization Strategy Facilitate Learning Analytics Design?* Paper presented at the the 27th International Conference on Computers in Education, Taiwan.
- Charleer, S., Klerkx, J., & Duval, E. (2014). Learning dashboards. *Journal of Learning Analytics*.
- Crawford, T., Kelly, G. J., & Brown, C. (2000). Ways of knowing beyond facts and laws of science: An ethnographic investigation of student engagement in scientific practices. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 37(3), 237-258.
- Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), *Collaborative learning: Cognitive and Computational Approaches* (pp. 1-19): Oxford: Elsevier.
- Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education*, 69, 485-492.
- Dillenbourg, P., & Jermann, P. (2010). Technology for classroom orchestration. In *New science of learning* (pp. 525-552): Springer.
- Dragan, G., Shane, D., & Abelardo, P. (2016). How do we start? State and Directions of Learning Analytics Adoption. In: International Council for Open and Distance Education.
- Duran, L. B., & Duran, E. (2004). The 5E Instructional Model: A Learning Cycle Approach for Inquiry-Based Science Teaching. *Science Education Review*, 3(2), 49-58.
- Fischer, F., & Dillenbourg, P. (2006). *Challenges of orchestrating computer-supported collaborative learning*. Paper presented at the 87th annual meeting of the American Educational Research Association (AERA).
- Hämäläinen, R., Kiili, C., & Smith, B. E. (2017). Orchestrating 21st-century learning in higher education: A perspective on student voice. *British Journal of Educational Technology*, 48(5), 1106-1118.
- Kelly, N., Thompson, K., & Yeoman, P. (2015). Theory-led design of instruments and representations in learning analytics: Developing a novel tool for orchestration of online collaborative learning. *Journal of Learning Analytics*, 2(2), 14-43.

- Kwon, K., Shin, S., Brush, T. A., Glazewski, K. D., Edelberg, T., Park, S. J., . . . Alangari, H. (2018). Inquiry learning behaviors captured through screencasts in problem-based learning. *Interactive Learning Environments*, 26(6), 839-855.
- Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist*, 57(10), 1439-1459.
- Luis, P. P., Martina Holenko, D., Israel, G., Mahmoud, A., & Walid, B. (2011). Orchestrating technology enhanced learning: a literature review and a conceptual framework. *Int. J. of Technology Enhanced Learning*, 3(6). doi:10.1504/IJTEL.2011.045449
- Marshall, J. C., & Horton, R. M. (2011). The relationship of teacher-facilitated, inquiry-based instruction to student higher-order thinking. *School Science and Mathematics*, 111(3), 93- 101.
- Merriam Webster Inc. (Ed.) (2019).
- Oliveira, A. W., Boz, U., Broadwell, G. A., & Sadler, T. D. (2014). Student leadership in small group science inquiry. *Research in Science & Technological Education*, 32(3), 1-17.
- Phua, J. (2013). Collaborative Science Inquiry. 17-22.
- Podgorelec, V., & Kuhar, S. (2011). Taking advantage of education data: Advanced data analysis and reporting in virtual learning environments. *Elektronika ir Elektrotechnika*, 114(8), 111-116.
- Sacha, D., Sedlmair, M., Zhang, L., Lee, J. A., Peltonen, J., Weiskopf, D., . . . Keim, D. A. (2017). What you see is what you can change: Human-centered machine learning by interactive visualization. *Neurocomputing*, 268, 164-175.
- Savery, J. R. (2015). Overview of problem-based learning: Definitions and distinctions. *Essential readings in problem-based learning: Exploring and extending the legacy of Howard S. Barrows*, 9, 5-15.
- Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., . . . Dillenbourg, P. (2017). Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research. *IEEE Transactions on Learning Technologies*, 10(1), 30-41.
- Song, Y., Cao, J., Tam, V. W. L., & Looi, C. K. (2019). *M-Orchestrate Learning System - A platform for orchestrating collaborative science inquiry*. Paper presented at the the 23rd Global Chinese Conference on Computers in Education (GCCCE2019).
- van Leeuwen, A. (2015). Learning analytics to support teachers during synchronous CSCL: Balancing between overview and overload. *Journal of Learning Analytics*, 2(2), 138-162.
- Watts, M. (2003). The orchestration of learning and teaching methods in science education. *Canadian Journal of Math, Science & Technology Education*, 3(4), 451-464.
- Webb, N. M. (2009). The teacher's role in promoting collaborative dialogue in the classroom. *British Journal of Educational Psychology*, 79(1), 1-28. doi:10.1348/000709908X380772

Understanding Jump Back Behaviors in E-book System

Boxuan Ma, Jiadong Chen, Chenhao Li,
Likun Liu, Min Lu, Yuta Taniguchi, Shin'ichi Konomi
Kyushu University, Japan
ma.boxuan.611@s.kyushu-u.ac.jp

ABSTRACT: With the increase of digital learning materials in higher education systems, a better understanding of student reading behavior and their effect on student performance get attention. Our research shows that, on average, each e-book system user uses “jump-back” to navigate a course slides for 12.7 times. In this paper, we aim to understand the student’s intention for a jump-back. We first formally define the problem of “jump-back” behaviors of reading slide at a face to face lecture, then we systematically study the problem from different perspectives on a real e-Book event stream data. Our study on the dataset reveals several interesting phenomena, e.g. students have different jump-back preferences. Also, students with a higher quiz score were having diverse jump-back behaviors, whereas the students with a comparably low quiz score are feasible to have a comparatively lower jump back frequency.

Keywords: reading behavior; e-Book event stream; educational big data; jump-back.

1 INTRODUCTION

Recently, Learning Management Systems (LMSs) and e-book systems are increasingly used together for supporting daily classroom teaching in many schools. These systems enable us to analyze the log data corresponding to students’ learning activities. Such activity log data represent one of the most valuable sources of information for analyzing the activities of students. Analyzing such data provides a novel and great potential for understanding students’ behaviors and enhancing education delivery. For example, using clickstream data to predict student performance [Brinton & Chiang, 2015; Okubo et al., 2017], to predict the class completion [Crossley et al., 2016] and clustering learner behaviors [Wang et al., 2016].

Event stream data from e-Book systems have been also utilized to understand students’ learning activities. Reading learning materials probably is the most important activity in current college education systems. Actually, the majority of the time that students spend on class is reading slides. Recently, researches have been conducted on the interactions between users and the e-book systems to better understand how students learn and what they need when reading learning materials. For example, pattern mining of preview and review activities [Oi et al., 2015], understanding learning behavior of students [Yin et al., 2015], browsing pattern mining [Shimada, Okubo & Ogata, 2016], and analysis of highlighters on e-textbooks [Taniguchi et al., 2019], etc. However, we found that the jump-back is a frequent behavior with strong user intention. Our preliminary study shows that, on average, each e-book system user uses “jump-back” 12.7 times to navigate a course slide. The reasons may include there is some difficult part that the student cannot understand and the student simply missed some part for other reasons.

In this paper, we conduct a systematic study as a first step to look into this problem in classroom setting by using students' reading logs that were collected from a digital textbook reader in order to better understand student reading behaviors.

The remainder of this paper is organized as follows. In Section 2, we describe the datasets that are used in this paper, then we introduce how we preprocess and analyze the dataset. In Section 3, we conduct the experiments to analyze jump-back behaviors from different perspectives, the details and results also are shown in this section. Finally, we will draw the conclusion and describe future work in Section 4.

2 METHOD

2.1 Data

As the data source, we used reading logs collected from a 90-minutes long in-class activity. Each student used the digital textbook reader (BookRoll) during the lecture. BookRoll is a system that allows digital materials to be delivered in lectures [Ogata et al., 2015; Ogata et al., 2017; Flanagan & Ogata, 2017]. Students can browse anytime and anywhere from a web browser on their personal devices (computer or smartphone). In the BookRoll system, there are features like highlighting, marking, memo function, etc. that students can use for learning. All click-stream were recorded in a database that is related to students' interaction with the system. At the end of the lecture, students took part in the quiz session related to content.

The collected click-stream data contained the following fields: *userid* (anonymized student user id), *contentsid* (the id of the e-book that is being read), *operationname* (the action that was done, e.g. open, close, next, previous, jump, add marker, add bookmark, etc.), *pageno* (the current page where the action was performed), *marker* (the reason for the marker added to a page, e.g. important, difficult), *memo_length* (the length of the memo that was written on the page), *devicecode* (type of device used to view BookRoll, e.g. mobile, pc), and *eventtime* (the timestamp of when the event occurred).

Table 1: Description of the Event Stream Dataset

Category	Type	Number
Lecture	Lecture Time	90 min
	Page Length	83
Student	Total Student #	118
Operation event	Total Event #	263,286
	Total PAGE JUMP #	7,087
	Total SEARCH JUMP #	71
	Total BOOKMARK JUMP #	1,559
	Total NEXT #	154, 401
	Total PREV #	70,360

There are different operations related to our research, i.e. PREV, NEXT, SEARCH JUMP, BOOKMARK JUMP, and PAGE JUMP. PREV means that the student clicked the PREV button to move to the previous page, and NEXT means that the student clicked the next button to move to the subsequent page. Students can also use PAGE JUMP/SEARCH JUMP/ BOOKMARK JUMP function to jump to another page. Table 1 lists the statistics of the event stream dataset for a specific lecture. We found that PAGE JUMP/SEARCH JUMP/ BOOKMARK JUMP operation is rare in the dataset, instead, students usually click the NEXT or PREV button quickly to jump to the desired page. For example, a student is on page 10 now and he/she wants to jump to page 5, then he would like to click the PREV button 5 times quickly instead of using PAGE JUMP function. To deal with such a problem, we introduced our method in the next section.

2.2 Data Preprocessing

Now we introduce how to preprocess the dataset to deal with the problem above. To start with, we first give the definition of the concept **Complete-jump (Cj)** and other events of which a complete-jump consists.

Definition 1. Complete-jump (Cj): A complete-jump consists of one (or multiple) jump-back actions by a specific student on a specific lecture slide, trying to find the right page to review. Let (s, l, ps, pe) denote a complete-jump, which means student s jumps back from start page ps to end page pe in slide l .

Definition 2. Jumping back (Jb): When a student use PAGE JUMP/SEARCH JUMP/ BOOKMARK JUMP function from the current page (ps) to jump back to another page (pe) ($pe < ps$) or click PREV button to go to the previous page, then we say there is a jumping back event.

We noticed that a complete-jump might consist of more than one jump action. For example, the student clicks the PREV button several times to jump back to a previous page. Another example, the student may jump back to a page of no interest and continue to look for the right page that she/he desires to review.

Definition 3. Jumping forward (Jf): When a student use PAGE JUMP/SEARCH JUMP/ BOOKMARK JUMP function from the current page (ps) to jump to a page afterward (pe) ($pe > ps$), or click NEXT button to go to the next page, then we say there is a jumping forward event.

There also might be jump-forward actions in a complete-jump. For example, the student jumps back far away from the desired page and then she/he jumps forward to adjust to the right position.

Definition 4. Short watching (Sw): After jumping to the desired page in the slide, the student usually would take a look for seconds. We name it as a short watching event. We use the short watching event to determine the end of a complete-jump. There is a duration period between two jumping events, i.e., from the time the first jumping event ends ($t1$) and the time the next jumping event occurs ($t2$). The duration period should be no longer than Sw , i.e., $t2 - t1 \leq Sw$. In our experiments, we tentatively set $Sw = 2$ seconds¹.

¹ We tried different settings for Sw and empirically selected 2s as an optimal setting.

Actually, complete-jump behavior cannot be obtained straightforwardly. Enlighted by [Zhang et al., 2017], we modified their algorithm based on deterministic finite automaton to reconstruct them.

Based on the definitions above, we use a deterministic finite automaton (DFA) to construct the complete-jump behaviors. Figure 1 shows the state transition in the DFA. There are four states: Ready, Record, Check, Dump. At the Ready state, it stays until receives a jumping back event (Jb), then the state goes to Record. When the state is Record, it maintains a stack. When there are jumping back events (Jb) or jumping forward (Jf) events, it pushes all the events into the stack. If there comes a short watching event (Sw) or some other operations (e.g., the student use MAKER or MEMO function), the state transforms to Check state. When the state is Check, it compares the start page (ps) of the event at the bottom of the stack and the end page (pe) of the event at the top of the stack. If $pe > ps$, the sequence of events in the stack constitutes a jump-forward behavior, then the state goes back to Ready. Otherwise, the state transforms to Dump, where we aggregate the sequence of events in the stack to construct a complete-jump behavior.

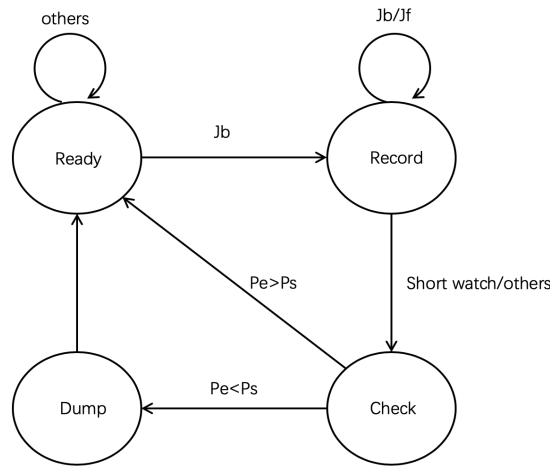


Figure 1: The construction of complete-jump behavior based on DFA

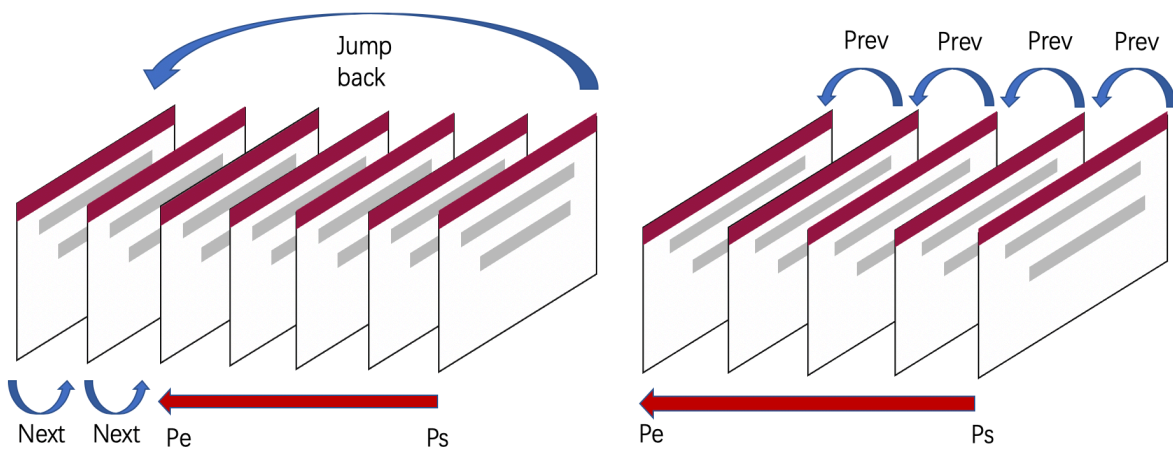


Figure 2: Two complete-jump patterns

Figure 2 shows two common complete-jump patterns in the dataset. The right pattern illustrates a kind of complete-jumps that consist of the event sequence $[Jb, Jb, Jb, Jb]$, which means that the student uses the PREV button 4 times to jump back to a previous page (pe). The left pattern shows a complete-jump that consists of the event sequence $[Jb, Jf, Jf]$. In this kind of scenario, the student uses PAGE JUMP operation to jump back to a previous page firstly and then clicks the NEXT button 2 times to jump to a later page (pe).

2.3 Data Analysis

For the data analysis, first we visualized all students' page flip patterns. Later, we engaged in some investigation of the complete-jump behavior of the students.

The investigations are conducted from three perspectives:

- (1) General performances: What is the general performance of students' jump back behaviors? How does the general performance vary in different lectures and slides?
- (2) Student preferences: Do students have personal preferences when they jump back? how they show their preferences?
- (3) Student Academic Performance: Are there any relationships between students' jump back behaviors and their academic performance?

We analyzed our data by employing basic statistical analysis methods as well as the k-means clustering algorithm to answer the questions above, the details and results will be described in the next section.

3 RESULT

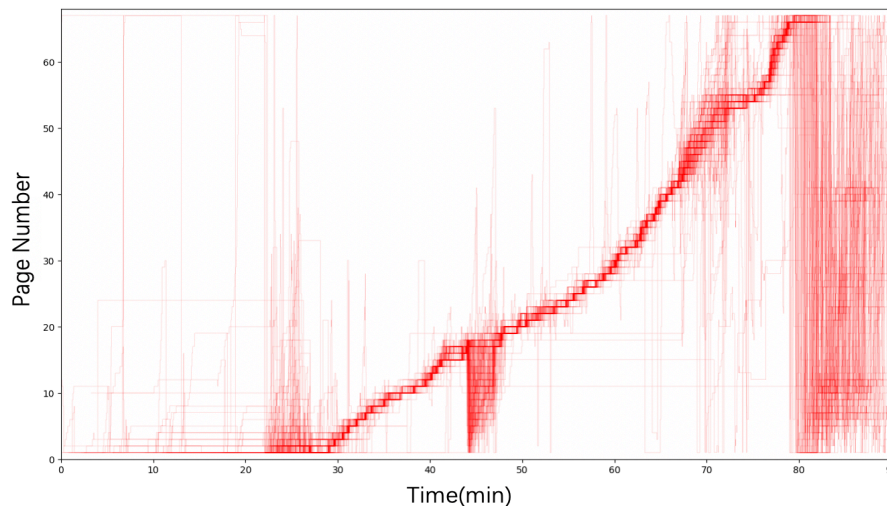


Figure 3: Students' page flip patterns across the lectures

Visualization of page flip patterns of all students can be seen in Figure 3. The X-axis shows the time, Y-axis shows the page of the slides. The intersection of the Time and Page shows the current page of the student at a specific time. Each line shows the reading patterns of a particular student. We can see that many students would like to take a quick look at the entire content in the first 20 minutes of the class, and they review the slide in the last 10 minutes of

the class. Figure 3 shows that jump-back is a frequent behavior with strong user intention. Then we will discuss our investigation results of the complete-jump behavior of a student in the following three subsections.

3.1 General Performance

To have a better understanding of students' general jump back performance in a lecture, we plot all the complete-jumps of the slide of a specific lecture in Figure 4. A spot (x, y) represents a complete-jump from the start page x to end page y . The figure shows that most spots are near the diagonal. It indicates that students usually do not jump back to a more distant page from the current page. We name the number of pages between the start page and end page as **jump span**. In this case, the jump span of 80% complete-jumps is smaller than 6 pages, shown as the red area. This phenomenon also exists in other lectures of the dataset.

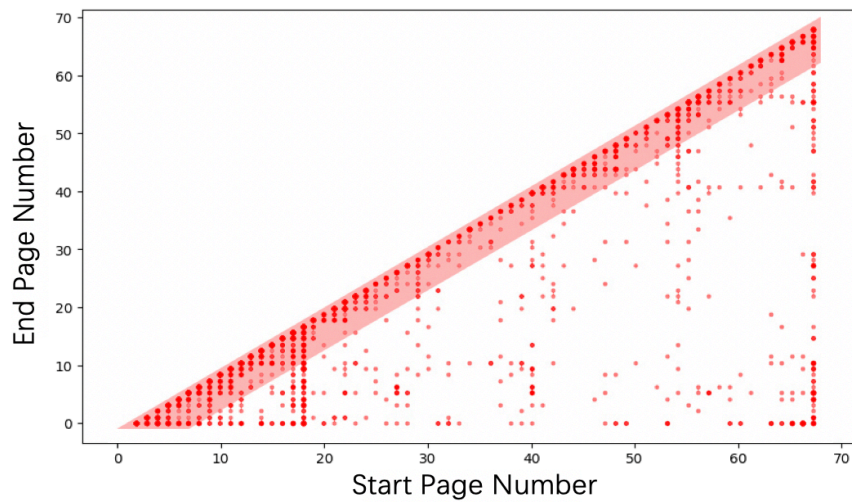


Figure 4: The scatter of complete-jumps. A spot at (x, y) represents a complete-jump from page x to page y .

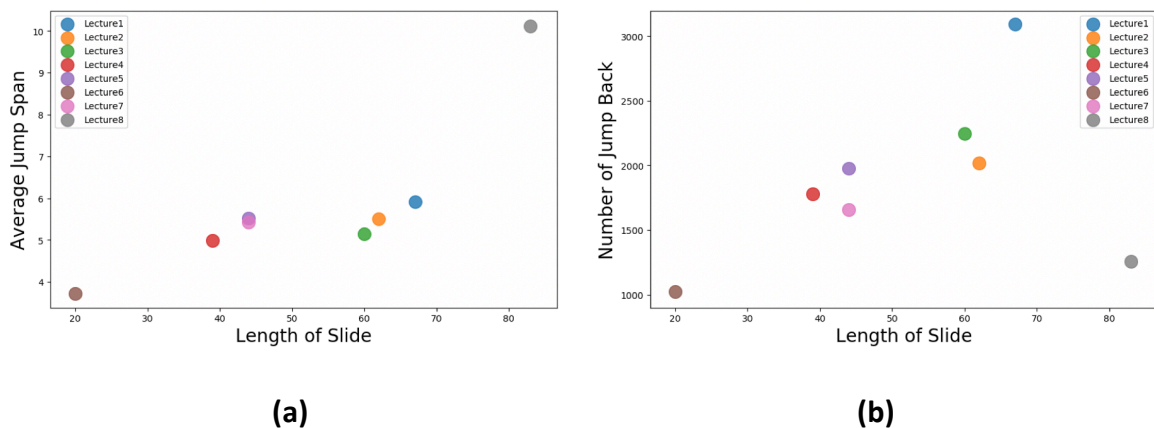


Figure 5: General complete-jump performance comparison in different lectures. Y-axis: (a) average jump span of each lecture, (b) number of complete-jumps of each lecture. X-axis: the lengths of slides for different lectures.

We have 8 lectures in the dataset. The lengths of slides for different lectures vary from 20 pages to 83 pages. We want to know whether the length of a slide has an effect on complete-jump behavior. Figure 5 (a) shows the correlation between slide length and jump span, and Figure 5 (b) shows the correlation between slide length and the number of complete-jump. The results show that the jump span and the complete-jump number is positively correlated with the length of slides. Where the abnormal occurred in lecture 8 of Figure 5 (b) is conceivable due to the fact that lecture 8 was the last lecture and students need to take the quiz.

3.2 Student Preference

Different students would have different jump-back patterns. For example, impatient students are likely to jump with higher frequency than patient students. To catch students' preferences, we categorize students into different types based on their jump back behaviors leveraging k-means clustering. Table 2 shows the clustering results. In Table 2, Average Stay Time indicates that the average time of reading after the student jumped back to the page.

Table 2: Clustering results of students' jump back records.

Item	C1	C2	C3
# of Jump Back	18	8.5	11.6
Max Jump Span	35	28	32
Min Jump Span	1	5	1.1
Average Jump Span	6	9.7	7.3
Average Stay Time(s)	39	292	101

Students of Clustering 1 have clear preference when they jump back, they prefer to jump more times (18 times) with short jump span (6 pages) and they stay short time after jumping to their desired page (39 seconds), while students of cluster 2 have a preference to jump back farther away (9.7 pages) with lower frequency (8.5 times), but they prefer to stay longer after jumping back to their desired page to have a serious reading (292 seconds). Students of Clustering 3 seem to have no obvious preference and their jump back behavior is more or less unpredictable.

3.3 Student Academic Performance

Table 3: Partial correlation results

Item		# of Jump Back	Average Jump Span	Average Stay Time
Quiz Score	PCC	0.1229	0.2422	0.0233
	P-value	0.1848	0.0082	0.2646

As mentioned before, students took the quiz of the last lecture. We use the Pearson correlation coefficient (PCC) to calculate the partial correlation of quiz scores with other variables, such as the number of the jump back and jump span, etc. Table 3 presents the

results of the partial correlation analysis. We can see that there is no significant correlation between the jump back behaviors and the quiz score.

Table 4: Comparison results between the high-score group (G1) and low-score group (G2)

Item	# of Jump Back	Average Jump Span	Average Stay Time (s)
G1 Mean (Std)	12.07(6.34)	6.29(3.17)	133.26 (165.58)
G2 Mean (Std)	10.53(5.02)	4.96(2.79)	101.30 (91.15)

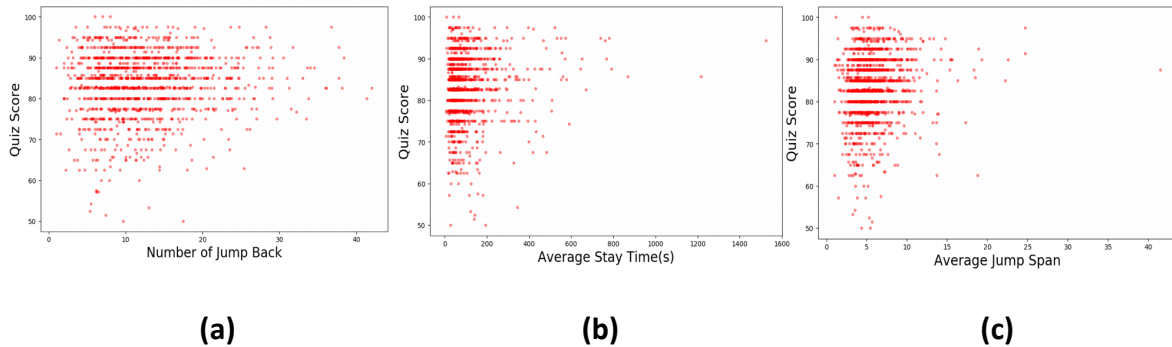


Figure 6: The scatter of complete-jump behavior and quiz scores. X-axis: (a) the average number of complete-jumps for each student, (b) average stay time for each student, (c) average jump span for each student. Y-axis: the quiz score.

To provide a clear comparison, we plot scatters in Figure 6 of (a) the average number of complete-jumps for each student and quiz score, (b) average stay time for each student and quiz score, (c) average jump span for each student and quiz score. We also split students into the high-score group (G1, score over 90, N=201) and low-score group (G2, score below 70, N=84) to compare the difference of jump back related features in Table 4. Based on the results above, a safe conclusion could be drawn that while the jump back behaviors vary among the students who have a relatively better quiz score, the students with a lower quiz score tend to have a lower frequency of jump back, shorter jump span and stay time.

4 CONCLUSION

This research aims to tackle the reading behavior of learners while using the e-book system to better understand how students read and learn. Particularly, this paper studied the student intention for a jump-back behavior. Through the analytics of e-Book event stream data, we first formally define the problem of “jump-back” behaviors of reading slide at the face to face lecture, then we systematically studied the jump-back behaviors from different perspectives. Our result shows several interesting phenomena, e.g. different students have different jump-back preferences. Students with a higher quiz score were having diverse jump back behaviors, whereas students with a comparably lower quiz score are feasible to have a lower jump back frequency.

As the data source of the current work is limited to page navigation events, it can be difficult to extract a more comprehensive understanding of the jump back behaviors. In our future work, taking other types of the students' behaviors and the teaching processes into account will be helpful to construct explanatory models of irregular page-viewing behaviors. Furthermore, it will invoke new approaches of feedback to the instructors and students, for example, suggestions of useful jump-back destinations of the current page, which will contribute to the improvement of learning efficiency.

REFERENCES

- Brinton, C. G., & Chiang, M. (2015). MOOC performance prediction via clickstream data and social learning networks. In *2015 IEEE conference on computer communications (INFOCOM)* (pp. 2299-2307). IEEE.
- Crossley, S., Paquette, L., Dascalu, M., McNamara, D. S., & Baker, R. S. (2016). Combining clickstream data with NLP tools to better understand MOOC completion. In *Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 6-14). ACM.
- Flanagan, B., & Ogata, H. (2017). Integration of learning analytics research and production systems while protecting privacy. In *The 25th International Conference on Computers in Education, Christchurch, New Zealand* (pp. 333-338).
- Oi, M., Okubo, F., Shimada, A., Yin, C., & Ogata, H. (2015). Analysis of preview and review patterns in undergraduates' e-book logs. In *23rd International Conference on Computers in Education, ICCE 2015* (pp. 166-171). Asia-Pacific Society for Computers in Education.
- Ogata, H., Oi, M., Mohri, K., Okubo, F., Shimada, A., Yamada, M., ... & Hirokawa, S. (2017). Learning analytics for e-book-based educational big data in higher education. In *Smart Sensors at the IoT Frontier* (pp. 327-350). Springer, Cham.
- Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-Book-based learning analytics in university education. In *International Conference on Computer in Education (ICCE 2015)* (pp. 401-406).
- Okubo, F., Yamashita, T., Shimada, A., & Ogata, H. (2017). A neural network approach for students' performance prediction. In *LAK* (pp. 598-599).
- Shimada, A., Okubo, F., & Ogata, H. (2016). Browsing-Pattern Mining from e-Book Logs with Non-negative Matrix Factorization. In *EDM* (pp. 636-637).
- Taniguchi, Y., Shimada, A., Yamada, M., & Konomi, S. I. (2019). Recommending Highlights on Students' E-Textbooks. In *Society for Information Technology & Teacher Education International Conference* (pp. 874-880). Association for the Advancement of Computing in Education (AACE).
- Wang, G., Zhang, X., Tang, S., Zheng, H., & Zhao, B. Y. (2016). Unsupervised clickstream clustering for user behavior analysis. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 225-236). ACM.
- Yin, C., Okubo, F., Shimada, A., Oi, M., Hirokawa, S., & Ogata, H. (2015). Identifying and analyzing the learning behaviors of students using e-books. In *23rd International Conference on Computers in Education, ICCE 2015* (pp. 118-120). Asia-Pacific Society for Computers in Education.
- Zhang, H., Sun, M., Wang, X., Song, Z., Tang, J., & Sun, J. (2017). Smart jump: Automated navigation suggestion for videos in moocs. In *Proceedings of the 26th international conference on world wide web companion* (pp. 331-339). International World Wide Web Conferences Steering Committee.

Social Knowledge Mapping Tool for Interactive Visualization of Learners' Knowledge

Akira Onoue¹, Masanori Yamada², Atsushi Shimada³, Tsubasa Minematsu⁴,
Rin-ichiro Taniguchi⁵

Kyushu University

onoue@limu.ait.kyushu-u.ac.jp¹, mark@mark-lab.net²,
{atsushi³, minematsu⁴, rin⁵}@ait.kyushu-u.ac.jp

ABSTRACT: The purpose of our study is to create a tool that makes it easy for teachers to analyze learners' knowledge maps and grasp useful information for conducting educational activities. Learners create their own knowledge maps to reflect their learning activities. Our system collects individual knowledge maps from learners and produces an integrated knowledge map: Social Knowledge Map (SKM). SKM tool enables interactive visualization and analysis of knowledge maps. We surveyed a professor of Instructional Science and learning supporters studying educational technology to evaluate the effectiveness of the tool. We received positive responses and realized the need for more analytical capabilities and collaboration with other types of learning logs.

Keywords: knowledge map, interactive visualization tool, use and evaluation

1 INTRODUCTION

By using digital learning materials, learner's reading logs can be collected. Based on reading behavior, clarifying how learners have organized the knowledge learned in lectures plays an important role in improving learning performance. A knowledge map is an effective cognitive learning tool. A knowledge map focuses on building knowledge relationships, such as links between information, thus creating subject/content-specific maps (Crampes et al, 2006) (Balaid et al, 2016). Lee and Segev (2012) showed that knowledge maps are effective in promoting idea generation. This suggests that a knowledge map is effective at granting access to knowledge in a timely manner, identifying knowledge flow, allowing organizational restructuring, and so on. Considering the effect of the knowledge map, it is considered possible to create a powerful tool to investigate the understanding process by integrating a mapping tool and an input-based system. This study aims to develop an educational support system, Social Knowledge Map (SKM), which can aggregate learners' knowledge maps from various perspectives. SKM shows how course contents are organized and remembered by learners. The information is important for both teachers and learners as it reflects their teaching and learning activities, respectively. In an actual educational situation, teachers cannot spend much time to check the knowledge maps of many learners. Therefore, we implement a tool that visualizes the integrated knowledge map created by combining each learner's knowledge map. A professor of Instructional Science and learning supporters studying educational technology evaluated the effectiveness of the tool.

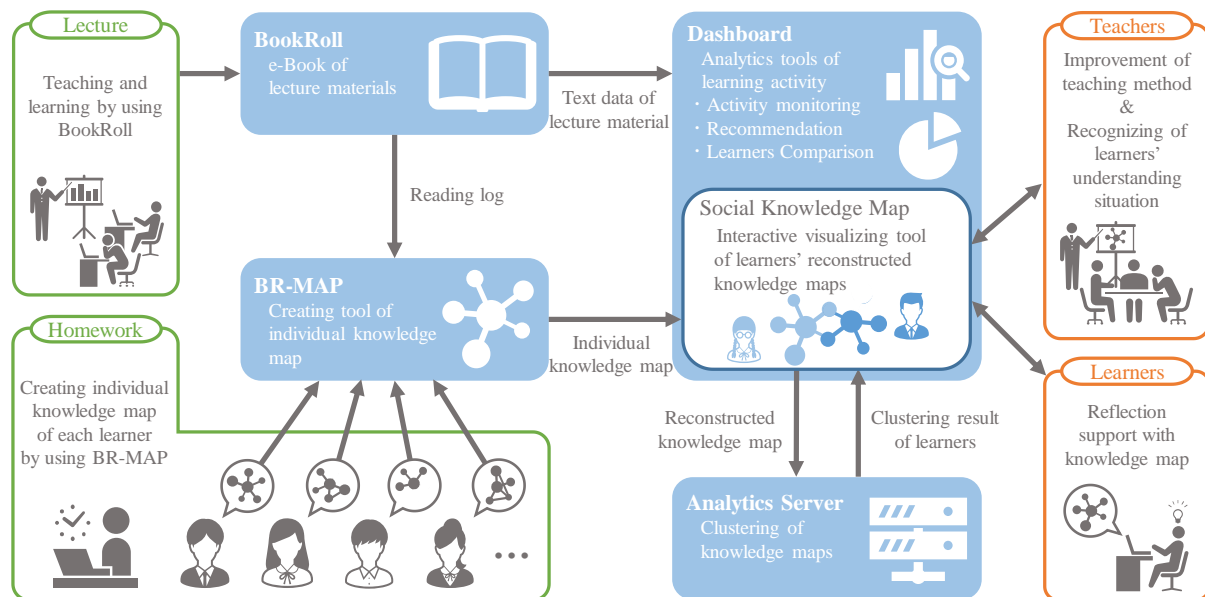


Figure 1: System Overview

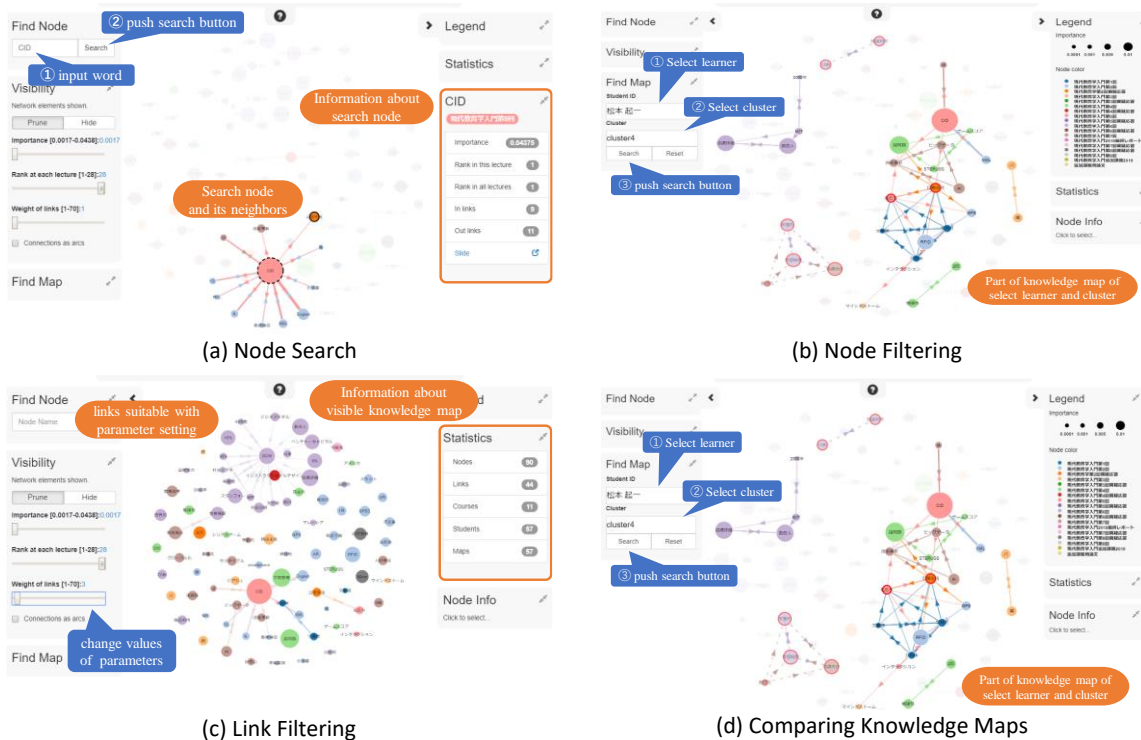
2 KNOWLEDGE MAP OF LEARNERS

A knowledge map is a network in which learned keywords are arranged as nodes and the relationships between the nodes are indicated using arrows. This section discusses an overview of SKM tool and how to create learners' knowledge maps. SKM tool is a knowledge map tool based on integrating individual knowledge maps created by learners using BR-MAP tool. The overview of the system configuration is displayed in Figure 1. The system consists of an e-Book viewer "BookRoll" (Ogata et al, 2017), the creating tool of each learners' individual knowledge map "BR-MAP" (Yamada et al, 2018), Dashboard System, and analytics server. The procedure of system usage is as follows: (1) Teachers conduct lectures using the BookRoll system. Learners open the lecture materials on the BookRoll and highlight or attach memos on the lecture materials. (2) Each learner creates one's knowledge map by using the BR-MAP system to reflect his or her learning activities. (3) On the Dashboard system, text data of lecture materials are analyzed to extract words suitable for SKM. (4) Reconstructing the original knowledge maps created in step (2). (5) Clustering learners based on their knowledge maps. (6) Using the tool to evaluate the level of understanding achieved by learners of the lectures from the resultant knowledge map visualization.

In step (1), (2), (3), and (4), our system analyzes the e-Book and the nodes/links to facilitate the integration of individual knowledge maps proposed by (Onoue et al, 2019a) (refer to (Onoue et al, 2019a) for details about integrating the learners' knowledge maps). In step (5), We applied a graph clustering method proposed in (Onoue et al, 2019b) to the reconstructed knowledge maps (refer to (Onoue et al, 2019b) for details about the process). We call the knowledge maps of learners included in each cluster's "sub-map".

3 SOCIAL KNOWLEDGE MAP TOOL

This section describes SKM tool that enables interactive visualization and comparison of knowledge maps. We used a method to create SKM based on (Onoue et al, 2019a). SKM tool enables learners and teachers to visualize, analyze, and compare knowledge maps on a web browser. We implemented

**Figure 2: Functions of SKM**

the functions of node search, node and link filtering, and comparison of knowledge maps in the proposed tool. Corresponding to the above functions, statistical information of the displayed knowledge map and detailed information of the node are presented. A larger-sized node represented an important node, which meant that many learners drew links to/from the node. Additionally, the nodes were color-coded to correspond with lectures that had learning materials in which the word was frequently used. We explain how to use this tool and the details of each function as follows: First, the system user selects a course or lectures for which a knowledge map is to be created. The user can select the entire course or only a part of the course. By allowing multiple data selection patterns, it is expected that data for creating knowledge maps can be used for various purposes.

As the knowledge maps of many learners are integrated and visualized, it may be harder to investigate the relationship between the focused node (knowledge) and other nodes. Therefore, implementing the node search function enables users to find the focused node quickly. An example of searching is shown in Figure 2(a). The user can use this function by entering a word in the input form of the "Node Search" tab and pressing the search button. When there is a word (node) that matches the input word in the knowledge map, that node and its neighbors are highlighted. In addition, the "Node Info" tab shows detailed information about the node of interest. The items are node name, lecture times that appeared most in lecture materials, the order of importance in all nodes, the order of importance in lecture times that appeared most in lecture materials, the number of incoming nodes, the number of outgoing nodes, and links to lecture materials.

According to (Onoue et al, 2019), there are differences in the size of the knowledge map considered optimal by users. In this study, we filter nodes and links displayed by adjusting three parameters to be able to provide information specifically for the user's intended purpose. The node parameters are importance and the ranking of importance in each lecture. The link parameter is the weight, that is,

the number of learners who have drawn links between nodes. The user changes the number of nodes/links displayed by adjusting the sliders for each parameter on the "Visibility" tab. When adjusting the parameters of nodes/links, only the nodes/links that satisfy the conditions and the links that connect with the nodes are displayed as showed in Figure 2(b), and 2(c). Since the three parameters can be used together, it is possible to adjust the scale of the knowledge map under multiple conditions. In addition, the "Statistics" tab shows detailed information about the knowledge map displayed. The items are the number of nodes, the number of links, the number of courses, the number of learners, and the number of individual knowledge maps.

This system visualizes not only the result of integration of learners' knowledge maps but also what part of SKM an individual learner's knowledge map constitutes. By comparing SKM with the learner's knowledge map, it is possible to confirm whether the learner understands important knowledge. We define three types of comparing between knowledge maps: (1) the learner's knowledge map and SKM, (2) sub-map and SKM, and (3) a learner's knowledge map and sub-map. In the "Find Map" tab, after selecting a student or/and a cluster, then pressing the search button, the knowledge maps of the selected learner or/and sub-map are highlighted is shown in Figure 2(d). We define the appearance of nodes/links based on the type of knowledge map to which the nodes/links belong. There are three types of nodes' appearance: A node that is not transparent at all and with no border is a node contained in a sub-map; a node that is not transparent at all and has a red border is a node contained in both a learner's knowledge map and a sub-map; and a translucent node with a red border is a node contained in only a learner's knowledge map. Additionally, there are two appearance types for links: A link with a solid line is a link contained in a sub-map and a link with a dotted line is a link contained in a learner's knowledge map.

4 EXPERIMENT

We conducted the experiments during the university education course to evaluate the SKM. The learners reviewed case studies on education and learning support systems. The goal of the course is to be able to explain an effective learning environment using ICT (Information and Communication Technology) based on these learning theories. The course was conducted over eight weeks from June to August 2019 at Kyushu University. In response to the question, "What is an effective learning environment using ICT?", we instructed learners to create a knowledge map using BR-MAP in the final lecture. In total, 33 first-year students created their individual knowledge map for this course.

We administered a questionnaire to the professor of Instructional Science for 20 years and to three learning supporters studying educational technology; one has been supporting learners for four years, and the others for two years. Participants were asked to evaluate this tool from the perspective of educational engineering. First, the participants read the manual of SKM tool. Next, after using SKM tool for 30 minutes on average, a questionnaire was administered as showed in Table 1.

We received positive responses to Q1 and Q2 from respondents. As for Q1, both teachers and learners expected to use this tool to improve educational and learning activities. Regarding the usage of this tool by teachers, we received responses that it is possible to improve the contents of lecture materials and review educational strategies by understanding learners' knowledge structures. In addition, using the knowledge maps, teachers can distinguish between students that have learned properly and those who have not. Thereafter, the tool can be used to encourage students to learn eagerly by showing

Table 1: The questionnaire about SKM

Question	Purpose
Q1 Is it effective to show SKM for improving classes and learning support?	To evaluate whether the system is useful for teachers in their educational activities
Q2 Is the comparison between different types of knowledge maps effectively in improving classes and learning support?	To evaluate whether educational activities could be improved by using the map comparison function
Q3 How would you design a course and use BR-MAP to make effective use of the SKM tool?	To reveal how BR-MAP should be used to create learners' knowledge maps for the purpose of using SKM effectively
Q4 What functions are necessary to effectively use this tool?	To clarify what functions of SKM should be implemented for respondents who actually work in educational fields
Q5 What kind of logs would you combine with SKM?	To reveal how respondents want to use other types of learning logs with SKM tool

knowledge maps of students who have learned properly to students who have not. By referring to this tool, it is possible for students to reflect connections between knowledge across multiple lectures. From answers to Q2, it is effective for learners to compare their own map with the other people's knowledge maps, because they can absorb the knowledge of others, and accordingly adjust and expand their own knowledge structures. It also effective for teachers: 1) By creating groups consisting of learners that belong to different clusters, and discussing the knowledge maps created within each group, a class applying the jigsaw method (Aronson et al, 1978) becomes possible. 2) Discovering the nodes that cause misunderstandings and misconceptions help teachers to improve classes and teaching materials.

Q3, Q4, and Q5 asked about the desired function of the SKM tool and improvements that could be made to the entire system. Further, we asked how using a future system with those improvements would affect the educational activities. Q3 revealed the following ideas about using BR-MAP: 1) Each learner creates a knowledge map as a reflection at the end of each lecture or as a lecture assignment. 2) They create a master knowledge map by adding new knowledge to their knowledge maps after every lecture. 3) A course manager adopts a lecture style that updates the knowledge map with each lecture. In order for SKMs to contribute towards improving learning activities and educational behavior, the rules for creating a knowledge map in BR-Map should be properly determined and informed students of the procedure for using the application beforehand.

The answers to Q4 mainly revealed that a function to visualize the knowledge map creation process in time series and a function to browse lecture materials is desired by respondents. Specifically, the change of the structure of an individual or SKM is visualized by operating the time (classroom) slider in the SKM tool. The answer to Q5 showed that reading logs of lecture materials can be used for the detailed explanation of each cluster. For example, if teachers can find which parts of lecture content learners in a particular cluster cannot understand, they can recommend references for a deeper understanding of individual topics and change homework based on the comprehension situation of learners in that cluster.

5 DISCUSSION AND CONCLUSION

We proposed SKM tool to enable interactive visualization and comparison of knowledge maps based on the e-Book contents and learners' knowledge map. The tool analyzes learners' knowledge maps and easily obtain useful information for conducting educational activities. To assess the effectiveness of the tool, we administered a questionnaire to a professor and learning supporters. We received positive responses about the usefulness of our tools. Two possible reasons emerged: First, respondents considered that it is possible to review the educational strategy by using the knowledge map to grasp a learner's knowledge structure throughout the course. Second, respondents considered that it is possible to propose different learning support methods for each learner by comparing knowledge maps of learners/clusters. There are four points that must be addressed in future research. First, we'll present the knowledge map of a teacher to the learner as a map representing the correct knowledge structure for the learner to be able to understand the course contents more easily. Second, we prioritize the implementation of a function that displays the time series of the knowledge map creation process. Third, by using other types of learning logs together, we enable teachers to interpret each learner's understanding more specifically and in detail. Fourth, we have to consider a better strategy to collect more sophisticated knowledge maps from students, because some students created their knowledge maps in an unexpected way. We will also administer a questionnaire to the learners to evaluate SKM tool from the learner's point of view.

ACKNOWLEDGEMENT

This work was supported by JST AIP Grant Number JPMJCR19U1, and JSPS KAKENHI Grand Number JP18H04125, Japan.

REFERENCES

- Aronson, E. (1978). *The jigsaw classroom*. Sage.
- Balaid, A., Rozan, M. Z. A., Hikmi, S. N., & Memon, J. (2016). Knowledge maps: A systematic literature review and directions for future research. *International Journal of Information Management*, 36(3), 451-475.
- Crampes, M., Ranwez, S., Villerd, J., Velickovski, F., Mooney, C., Emery, A., & Mille, N. (2006). Concept Maps for Designing Adaptive Knowledge Maps, *Information Visualization*, 5(3), 211-224.
- Lee, J.H., & Segev, A. (2012). Knowledge maps for e-learning, *Computers & Education*, 59, 353-364.
- Ogata, H., Oi, M., Mohri, K., Okubo, F., Shimada, A., Yamada, M., ... & Hirokawa, S. (2017). Learning analytics for e-book-based educational big data in higher education. In *Smart Sensors at the IoT Frontier*, 327-350. Springer, Cham.
- Onoue, A., Yamada, M., Shimada, A., & Taniguchi, R. I. (2019a, March). The Integrated Knowledge Map for Surveying Students' Learning. In *Society for Information Technology & Teacher Education International Conference*, 838-846. AACE.
- Onoue, A., Shimada, A., Minematsu, T., & Taniguchi, R. I. (2019b). Clustering of Learners Based On Knowledge maps. In 16th International Conference on Cognition and Exploratory Learning in the Digital Age, CELDA 2019. IADIS Press (to appear).
- Yamada, M., Shimada, A., Oi, M., Taniguchi, Y., & Konomi, S. (2018). BR-Map: Concept Map System Using E-Book Logs, *Proceedings of 15th International Conference of Cognition and Exploratory Learning in Digital Age*, 248-254.

Can the Area marked in eBook Readers Specify Learning Performance?

Xu Yufan¹, Geng Xuewang¹, Chen Li¹, Hamada Satomi¹,

Taniguchi Yuta², Ogata Hiroaki³, Shimada Atsushi⁴, Masanori Yamada^{1,5}

Graduate School of Human-Environment Studies, Kyushu University¹

Faculty of Information Science and Electrical Engineering, Department of Informatics²

Learning and Educational Technologies Research Unit Academic Center for Computing and
Media Studies, Kyoto University³

Faculty of Information Science and Electrical Engineering, Kyushu University⁴

Graduate School of Human-Environment Studies, Kyushu University⁵

xuyufan@mark-lab.net, geng@mark-lab.net, chenli@mark-lab.net, satomi.hamada@mark-
lab.net, Taniguchi.yuta.941@m.kyushu-u.ac.jp, hiroaki.ogata@gmail.com,
atsushi@ait.kyushu-u.ac.jp, mark@mark-lab.net

ABSTRACT: Learning behaviors are associated with learning performance. Using the “BookRoll” digital learning material readers to collect log data, researchers can analyze logs and teachers can provide feedback to learners. The marker functions in BookRoll help students focus on learning content for future tests based on cognitive and metacognitive strategies. In this study, learners were instructed to draw yellow and red markers on slide pages to indicate unfamiliar and important content, respectively. By analyzing this learning behavior, we show how it changes subjects’ learning performance. Studies suggest that the frequency of using markers during learning is associated with one’s learning performance. Data were collected from 80 senior high students; we compared correlations between learning performance, frequency of the marker, and the area of marker. Mann-Whitney-U-Test results showed that students in ascending groups were ranked higher than those in the descending group based on the frequency of red markers, whereas students in ascending groups were ranked lower based on the frequency of yellow markers, as well as the area of red and yellow markers. The finding also indicates that the area marked is more strongly correlated with learning performance than the frequency of using markers.

Keywords: Knowledge Monitoring, Learning Analytics, Learning Behaviors, Learning Performance

1 INTRODUCTION

Learning behaviors exert an influence on learning performance (Janssen & O’Brien, 2014; Yamada et al., 2017). These behaviors appear in different self-regulated learning (SRL) phases. A self-regulated learner can metacognitively, motivationally, and behaviorally adjust to and master his study (Zimmerman & Schunk, 1989). SRL is believed to be associated with learning performance (Wang, 2011). With the improvement of Information and Communication Technology (ICT) in education, these learning processes can be recorded as learning logs by Learning Management Systems (LMS) such as Moodle (Ogata et al., 2015, Ogata et al. 2017). The Learning Analytics (LA) approach is used to analyze these logs. For instance, learners can use digital markers to highlight what they have read, and these logs can be analyzed using LA to specify their correlation with learning performance. Van Horne et al. (2016) demonstrate that using markers is a cognitive learning strategy and improves

comprehension. Yamada et al. (2017) found that marker functions are strongly correlated to the enhancement of SRL awareness. While markers are considered to be significant tools in learning, overusing or misusing them affects learning outcomes negatively. Studies have focused largely on the frequency of using markers without considering the area marked as other factors (Yin et al., 2019; Al-khazraji, 2019). In this study, we aim to investigate whether the effects of the area marked has a stronger correlation to learning performance as compared to the frequency of using markers.

2 LITERATURE REVIEW

2.1 SRL in Computer-Based Learning and Learning Analytics

SRL is a learning process that involves cognitive and metacognitive strategies, motivation, and behavior, all of which have an influence on learning performance (Pintrich et al., 1990). Cognitive strategies include awareness, feeling, prediction, and evaluation, whereas metacognitive strategies include goal-setting, planning, monitoring, and revision (Tobias & Everson, 2002). High achieving learners who have good metacognitive knowledge use this skill to arrange their time wisely, set goals, and self-reflect. On the other hand, learners who have poor metacognitive knowledge receive low scores on learning performance and are unconfident about their cognitive strategies which have an influence on performance (Romainville, 1994). With the development of ICT, researchers are beginning to focus on the effect of SRL in reinforcing learners' knowledge. In addition, metacognition has become another focus in improving students' learning in computer-supported environments. For example, knowledge monitoring ability, which is known to be a metacognitive strategy that is strongly correlated to learning performance, can be improved by training agents in a computer-based learning environment (Kautzmann & Jaques, 2019). In a web-based learning environment, incorporation of SRL can inform learners about what is to be done, as well as how and when to do it (Kramarski & Michalsky, 2013). The function of markers in e-reader systems is recognized as cognitive and metacognitive strategy (Van Horne et al., 2016). Learners who have good metacognitive knowledge identify unfamiliar or important content and highlight it in an appropriate range, whereas those who have poor metacognitive knowledge do not know what is unfamiliar or important and overuse or misuse marker functions. In this study, using a computer-based learning environment, learning behaviors—especially marker functions related to cognition and metacognition—were recorded as log data, and the LA approach was applied to collect and analyze this learning log data and test score data. We hypothesize that learners who are aware of what they know use markers appropriately. This approach provides a new method to analyze the relationship between marker functions using cognitive and metacognitive strategies and learning performance.

2.2 Learning Analytics and Learning Behaviors on Markers

The LA approach enables us to record learning behaviors such as duration of reading electronic textbooks and specific content that was highlighted. Shimada et al. (2015) found that it is possible to collect data on learning behavior outside of class, and these data contribute to learning performance and can be collected as learning logs which can explore correlations between learning performance, learning strategies, learning skills, etc. Various collections of learning logs can be used to clarify different correlations and improve learning design and learning support. For example, by analyzing the frequency of using e-reader operations such as NEXT, PREV, and MARKER recorded by the BookRoll system, Nian et al. (2019) explored correlations between these operations and learning

performance and used machine learning to predict learning performance. Nian et al.'s (2019) study demonstrated that students who had high frequency of using markers were more inclined to achieve high learning performance. This finding supports instructors and researchers develop marker functions and offer advice on how often students should use markers. Yin et al. (2019) used e-reader log data to analyze the frequency of behavior patterns, such as deleting markers after adding markers or turning to the next page after turning to the previous page. The patterns of markers implied that teachers should highlight important content or provide hints for students who find it difficult to identify important content. Results of learning behavior pattern analysis help developers improve e-reader systems and aid researchers and learning designers with obtaining a deeper understanding of students' learning behaviors. By investigating functions of markers used by learners to indicate what they know and what is important, researchers have found that using markers largely influences learning performance (Al-khazraji, 2019). While the area marked also plays a role in learning performance, many researchers have drawn conclusions by merely analyzing the frequency of marker use (Yin et al., 2015; Yin et al, 2019; Nian et al., 2019).

With the development of LA, in addition to the frequency of using markers, the areas markers can be measured. In this study, two issues are addressed. First, by comparing the frequency of using markers with the area marked, changes in grades are specified (between intermediate exams and term-end exams). Second, rankings of frequency of marker use or area marked are compared between ascending and descending groups. In the ascending group, students belonged to a higher rank group in the term-end exam than the intermediate exam. While students who assigned to the descending groups belonged to a lower rank group in the term-end exam than the intermediate exam. Four research questions (RQ) were set:

RQ1: Between ascending and descending groups, which ranked higher based on frequency of yellow markers?

RQ2: Between ascending and descending groups, which ranked higher based on the total area of yellow markers?

RQ3: Between ascending and descending groups, which ranked higher based on the frequency of red markers?

RQ4: Between ascending and descending groups, which ranked higher based on the total area of red markers?

3 METHODOLOGY

This study measured online learning logs of markers over seven weeks (between the intermediate and term-end exam) and compared the frequency of marker use with the area marked in terms of their ability to specify changes in grades, which include ascending and descending groups. The study was conducted in Mathematics classes.

3.1 Participants and the course

The participants were 80 high school students in grade 10. Students could read slide pages uploaded by their teachers using a tablet device. The study was performed in Mathematics (one class/day)

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

classes. Students were divided into three Mathematics classes ($n(\text{class-A}) = 32$, $n(\text{class-B1}) = 24$, $n(\text{class-B2}) = 24$) based on their adeptness level which decided by their term-end exam last semester.

Before each class, teachers uploaded learning materials onto the BookRoll system (Ogata et al., 2017), where students could make notes and highlight content. Students were asked to preview these materials. In class, students were made to read 16 textbooks totally using the BookRoll system and follow instructions from teachers. Outside of class, students completed 13 assignments totally and reviewed their learning on the BookRoll system.

3.2 Data collection

Data were collected using two methods: tests score imitating learning performance and learning logs which represent learning behaviors. First, data were collected from the intermediate exam and term-end exam, which were marked out of 100 and conducted by the senior high school. Figure 1 shows the distribution of the intermediate exam and term-end exam scores with a kernel density estimate. Second, learning logs which depicted students' learning behaviors on BookRoll system were collected. The function of the marker is to clarify students' understanding (using the red marker) and important content was highlighted. We extracted logs of red and yellow markers from 113,538 logs of students' learning behaviors.

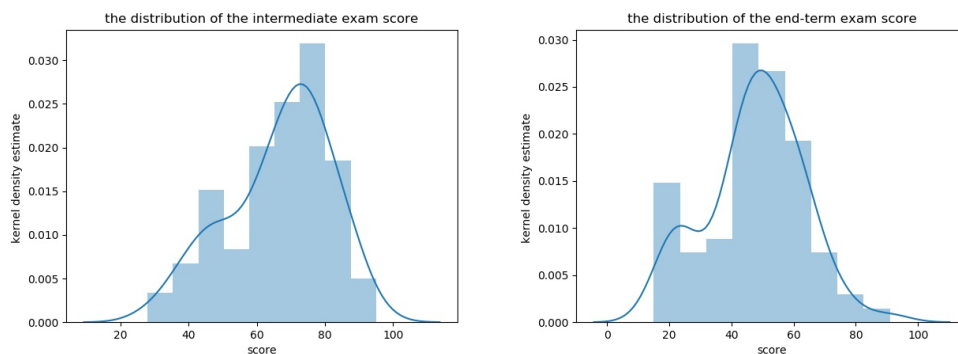


Figure 1: the distribution of the intermediate exam and term-end exam score with a kernel density estimate

3.3 Variables and data analysis

The Mann-Whitney-U-Test was used to measure the impact of the marker for learning behaviors on students' learning performance between the ascending and descending group. To assign these students into two groups, we calculated the increase or decrease between the intermediate exam and the term-end exam scores as follows:

3.3.1 Ascending and descending group

We assign each student into three groups, one standard deviation (1SD) away from the highest score, 1SD away from the lowest score, and others. Based on scores of the intermediate and term-end exam, each student was placed in one of three groups which were ranked high, medium, and low. Students in the ascending group belonged to a higher rank group in the term-end exam than

the intermediate exam. Students in the descending group were assigned to a lower ranking group in the term-end exam. Therefore, this study can validate the difference between these two groups.

To identify the marker variable, we calculated the frequency and the red and yellow areas marked as follows:

3.3.2 Frequency

F is the frequency of marker use by each student on the learning material system BookRoll. S_i represents a student in the group ($i=1,2,...,N$) and g represents the group ($g=1,2$). N may be different for different groups. j represents one marker. Therefore, for a single student in a group, the frequency of marker use can be quantified as:

$$F_{S_i}^g = \sum_{j=1}^M F_{S_{ij}}^g$$

Where each $F_{S_{ij}}^g$ represents 1.

3.3.3 Area

A represents the area of the marker, which the shape of a rectangle. The length and width of the rectangle were used to calculate the its area, representing the area of marker. The format of the marker's area in the log data has six dimensions (x, y, w, h, p_w, p_h), wherein x and y represent the vertical ordinate and horizontal ordinate; w and h represent the width and the height of marker; p_w and p_h are the width and height of the display device; the top left coordinate of the marker is (x, y); (p_w, p_h) is the slide display size, nevertheless this size is different from each display device. Therefore, the relative area calculated by dividing p_w, p_h by p_w, p_h for a single student in a group was as follows:

$$A_{S_i}^g = \frac{w}{p_w} \times \frac{h}{p_h}$$

Apparently, this value is less than 1. The total of relative area for a single student in a group was as follows:

$$A_{S_i}^g = \sum_{j=1}^R A_{S_{ij}}^g$$

4 RESULT AND DISCUSSION

To answer the fourth research question, we defined four different dimensions, frequency of the red marker (MARKER_RED_FRE), frequency of the yellow marker (MARKER_YEL_FRE), sum area of the red marker (MARKER_RED_AREA), and sum area of the yellow marker (MARKER_YEL_AREA) to represent the markers (See Table I). As the descending group and both samples (descending and ascending groups) were small, samples that tend to be larger are followed by four research questions. The non-parametric Mann-Whitney U test (one-tailed) was conducted on 3-mathematics-classes including class-A, class-B1, class-B2 in class (Table I) and out-of-class (Table II) between ascending group (M_THREE_H) and descending group (M_THREE_L), class-A mathematics in class (Table III) and out-of-class (Table IV) between ascending group (M_A_H) and descending group

(M_A_L), class-B1 mathematics in class (Table V) and out-of-class (Table VI) between ascending group (M_B1_H) and descending group (M_B1_L).

4.1 Descriptive Data and Mann-Whitney U Test Analysis of Variables

Table I presents the descriptive data of learning logs of markers for three-mathematics classes. This table shows M_THERE_L was larger on median than M_THERE_H, so the null hypothesis that two samples come from the same population and the alternative hypothesis that M_THERE_L tend to larger than M_THERE_H. The result of the Mann-Whitney U test (one-tailed) is that there is no statistical significance of difference between two groups. Alternatively, table II presents the descriptive data of learning log of marker for three-mathematics classes of students out-of-class. It represents that M_THERE_L is significantly larger for MARKER_YEL_FRE (U = 25.5, p = .023) and MARKER_YEL_AREA (U = 25.0, p = .020) than M_THERE_H. Furthermore, as for the red marker, M_THERE_L is larger on MARKER_RED_AREA (U = 31.5, p = .049).

TABLE I. Descriptive data for the three Mathematics class markers in class and Mann-Whitney U test (one-tailed) (n(M_THREE_L) = 24, n(M_THREE_H) = 5)

Learning log	Average (SD)		Median		U	p
	M_THREE_L	M_THREE_H	M_THREE_L	M_THREE_H		
MARKER_RED_FRE	17.83(25.76)	7.60(7.23)	5.00	6.00	49.5	.280
MARKER_YEL_FRE	16.88(27.43)	4.00(4.77)	5.00	1.00	50.0	.286
MARKER_RED_AREA	1.95(3.90)	0.99(1.40)	0.29	0.09	51.0	.311
MARKER_YEL_AREA	1.59(4.13)	0.71(0.94)	0.18	0.06	59.0	.488

***: p < 0.001, **: p < 0.01, *: p < 0.05, †: p < 0.1

TABLE II. Descriptive data for the three Mathematics class markers out-of-class and Mann-Whitney U test (one-tailed) (n(M_THREE_L) = 24, n(M_THREE_H) = 5)

Learning log	Average (SD)		Median		U	p
	M_THREE_L	M_THREE_H	M_THREE_L	M_THREE_H		
MARKER_RED_FRE	17.21(25.13)	2.00(3.10)	10.00	0	41.0	.138
MARKER_YEL_FRE	13.38(14.62)	0.20(0.40)	9.00	0	25.5*	.023
MARKER_RED_AREA	1.26(1.55)	0.20(0.37)	0.57	0	31.5*	.049
MARKER_YEL_AREA	1.44(1.75)	0.07(0.08)	0.79	0	25.0*	.020

***: p < 0.001, **: p < 0.01, *: p < 0.05, †: p < 0.1

Descriptive data of learning logs of marker in the Mathematics class-A of students in class is reported in Table III. Since M_A_L is less than M_A_H on the median, the alternative hypothesis is that M_A_L tend to be smaller than M_A_H. The statistical value indicated that M_A_H is significantly larger on MARKER_RED_FRE (U = 31.0, p = .027). Table IV reveals descriptive data of learning logs of markers for mathematics class-A of students out-of-class. Results of the Mann-Whitney U test (one-tailed) are that M_A_L is significantly larger on MARKER_YEL_FRE (U = 2.5, p = .009) and MARKER_YEL_AREA (U = 3.5, p = .014) and that M_A_L has a larger trend towards significance on MARKER_RED_AREA (U = 7.5, p = .052).

Table III Descriptive data for the Mathematics class-A of marker in class and Mann-Whitney U test (one-tailed) (n(M_A_L) = 9, n(M_A_H) = 4)

Learning log	Average (SD)	Median	U	p
--------------	--------------	--------	---	---

	M_A_L	M_A_H	M_A_L	M_A_H		
MARKER_RED_FRE	21.33(28.39)	121.00(121.96)	3.00	70.50	31.0	.027*
MARKER_YEL_FRE	17.67(26.26)	24.00(15.18)	7.00	27.50	25.0	.156
MARKER_RED_AREA	2.85(5.26)	6.21(3.67)	0.11	7.43	20.0	.408
MARKER_YEL_AREA	1.47(2.46)	2.28(1.67)	0.59	2.16	20.0	.408

***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, †: $p < 0.1$

Table IV Descriptive data for the Mathematics class-A of marker out-of-class and Mann-Whitney U test (one-tailed) ($n(M_A_L) = 9$, $n(M_A_H) = 4$)

Learning log	Average (SD)		Median		U	p
	M_A_L	M_A_H	M_A_L	M_A_H		
MARKER_RED_FRE	11.67(10.72)	15.00(25.98)	9.00	0	13.5	.258
MARKER_YEL_FRE	0.50(0.87)	17.44(15.09)	9.00	0	2.5**	.009
MARKER_RED_AREA	0.83(0.91)	0.02(0.03)	0.57	0	7.5†	.052
MARKER_YEL_AREA	2.09(2.18)	0.04(0.07)	1.37	0	3.5*	.014

***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, †: $p < 0.1$

Table V and Table VI express descriptive data for mathematics class-B1 of marker in class and out-of-class individually. Only M_B1_L has a larger trend towards significance on MARKER_YEL_AREA ($U = 13.5$, $p = .096$), as seen in Table V. And under Table VI, M_B1_L is significantly larger than M_B1_H on MARKER_YEL_AREA ($U = 10.5$, $p = .045$).

We drop tables expressing descriptive data for mathematics class-B1 of marker in class and out-of-class individually because of little logs of marker and no significance between two groups in the frequency of using marker and the total area of marker.

Table V Descriptive data for the Mathematics class-B1 of marker in class and Mann-Whitney U test (one-tailed) ($n(M_B1_L) = 8$, $n(M_B1_H) = 6$)

Learning log	Average (SD)		Median		U	p
	M_B1_L	M_B1_H	M_B1_L	M_B1_H		
MARKER_RED_FRE	21.50(24.40)	10.33(10.96)	17.50	8.50	15.5	.146
MARKER_YEL_FRE	17.63(23.00)	5.17(5.61)	8.50	3.00	14.5	.120
MARKER_RED_AREA	1.00(0.70)	1.05(1.31)	1.02	0.56	19.5	.301
MARKER_YEL_AREA	0.54(0.55)	0.49(0.87)	0.47	0.02	13.5†	.096

***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, †: $p < 0.1$

Table VI Descriptive data for the Mathematics class-B1 of marker out-of-class and Mann-Whitney U test (one-tailed) ($n(M_B1_L) = 8$, $n(M_B1_H) = 6$)

Learning log	Average (SD)		Median		U	p
	M_B1_L	M_B1_H	M_B1_L	M_B1_H		
MARKER_RED_FRE	11.88(12.05)	39.83(75.93)	6.00	1.00	19.0	.278
MARKER_YEL_FRE	17.44(15.09)	0.50(0.87)	9.00	0	17.5	.216
MARKER_RED_AREA	0.43(0.44)	0.63(0.84)	0.39	0.26	22.5	.448
MARKER_YEL_AREA	1.64(1.62)	0.49(0.75)	1.15	0.09	10.5*	.045

***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, †: $p < 0.1$

4.2 Effects of the frequency of using marker and the total area of marker on changes of grades between the ascending groups and the descending groups

RQ1: Between ascending and descending groups, which ranked higher based on the frequency of yellow markers?

RQ2: Between ascending and descending groups, which ranked higher based on total area of yellow markers?

Table II and Table IV indicate that descending groups are ranked higher than those in the ascending groups based on the frequency of using yellow markers (answer RQ1), whereas Table II, Table IV, Table V, Table VI indicates that descending groups were ranked higher than those in ascending groups based on the total area of yellow markers (answer RQ2). The knowledge monitoring ability, which is known to be a metacognitive strategy, involves checking the degree of understanding and affects learning performance (Hofer et al., 1998). In addition, studies show the importance of knowing what you know (Tobias & Everson, 2009). Learners who have poor metacognitive knowledge obtain low scores on their learning performance (Romainville, 1994). These learners would not know what they know, therefore they overuse and misuse the yellow markers and receive low learning performance, whereas learners who have good knowledge ability and know what they know use yellow markers appropriately and highlight using yellow markers in an appropriate range.

RQ3: Between ascending and descending groups, which ranked higher based on the frequency of red markers?

RQ4: Between ascending and descending groups, which ranked higher based on the total area of red markers?

Table III shows that descending groups rank lower than those in ascending groups based on the frequency of use of red markers (answer RQ3), whereas Table II and Table IV show that descending groups are larger than those in the ascending groups based on the total area of red markers (answer RQ4). The knowledge monitoring ability affects learning performance (Al-Harthy, 2011). Learners in ascending groups give an appropriate range of red markers, therefore their total area of red markers is lower than those in descending groups. Those who have good knowledge monitoring know important content. Therefore, learners in ascending groups have higher frequency of using red markers than those in descending groups. We draw the conclusion after comparing the significance on test score between two groups in class and out-of-class each class. Under teachers' tutoring In class, students can highlight content what is unfamiliar or important regardless of ascending groups or descending groups. On the other hand, without teachers' tutoring out-of-class, students in ascending groups who have good knowledge monitoring know what is unfamiliar or important, therefore, they use markers in an appropriate range.

5 CONCLUSION AND FUTURE WORK

In this paper, we analyzed learning logs of markers on the BookRoll reading system and compared the frequency of marker use with the area marked. The descending group is more inclined to use yellow markers and their frequency and area is larger than the ascending group, which means that

the descending group does not know what they know; conversely the ascending group is more inclined to use the red marker, and their frequency of use is higher than the descending group, which indicates that they know which is important. This finding also indicates that the area marked is more strongly correlated to learning performance than the frequency of marker use. The limitations of this study are the sample size of students and the type of content over which the student highlighted. The end-term test is more complicated than the intermediate exam, hence, the number of the ascending groups' students is small and the descending groups' students are large. We don't know which markers from textbook or an exercise sheet. This study implies that developers should take account of the area marked in reading system by learners. The implication of this study also shows that to these students in descending groups who don't have good knowledge monitoring need the teachers' tutoring out-of-class. Furthermore, the development of reading system should considerate the functions of the field of teacher education, digital learning material reader development or learning analytics.

In future work, we will use this finding to try to predict learning performance by using machine learning or deep learning using various data such as knowledge map construction data (e.g., Yamada et al, 2018). Similar to Okubo's research (2016), we could predict learning performance using other learning log data. Moreover, we could combine information of the markers with the dashboard and visualize it to improve learners' performance and implement the instructional design based on learning analytics such as Chen et al (2019).

ACKNOWLEDGEMENT

This research is partially supported by JST AIP Grant No. JPMJCR19U1, Japan

REFERENCES

- Al-Harthy, I. S. (2011). Knowledge monitoring, goal orientations, self-efficacy, and academic performance: A path analysis (Doctoral dissertation, Kent State University).
- Al-khazraji, A. (2019). Analysis of Discourse Markers in Essays Writing in ESL Classroom. *International Journal of Instruction*, 12(2), 559-572.
- Chen, L, Yoshimatsu, N., Goda, Y., Okubo, F., Taniguchi, Y., Oi, M., Konomi, S., Shimada, A., Ogata, H., & Yamada, M. (2019) Direction of collaborative problem solving-based STEM learning by learning analytics approach, *Research and Practice in Technology Enhanced Learning*, 14(1), 24.
- Hofer, B. K., Yu, S. L., & Pintrich, P. R. (1998). Teaching college students to be self-regulated learners. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practice* (p. 57–85). Guilford Publications.
- Janssen, S., & O'Brien, M. (2014). Disentangling the Effects of Student Attitudes and Behaviors on Academic Performance. *International Journal for the Scholarship of Teaching and Learning*, 8(2), n2.
- Kautzmann, T. R., & Jaques, P. A. (2019). Effects of adaptive training on metacognitive knowledge monitoring ability in computer-based learning. *Computers & Education*, 129, 92-105.
- Kramarski, B., & Michalsky, T. (2013). Student and teacher perspectives on IMPROVE self-regulation prompts in web-based learning. In *International handbook of metacognition and learning technologies* (pp. 35-51). New York: Springer.

- Nian, M. W., Lee, Y. H., & Wu, J. Y. (2019). Using machine learning to explore the associations among e-reader operations and their predictive validity of learning performance. *International Conference on Learning Analytics & Knowledge (LAK19)* (pp. 1-6).
- Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015, January). E-Book-based learning analytics in university education. *International Conference on Computer in Education (ICCE 2015)* (pp. 401-406).
- Ogata, H., Taniguchi, Y., Suehiro, D., Shimada, A., Oi, M., Okubo, F., Yamada, M., & Kojima, K. (2017). M2B system: A digital learning platform for traditional classrooms in university, Practitioner Track Proceedings of the 7th International Learning Analytics & Knowledge Conference, 154-161
- Ogata, H., Oi, M., Mohri, K., Okubo, F., Shimada, A., Yamada, M., Wang, J., & Hirokawa, S. (2017). Learning analytics for e-book-based educational big data in higher education. In Yasuura, H., Kyung, C. M., Liu, Y., Lin, Y. L. (Eds.), *Smart Sensors at the IoT Frontier* (pp. 327-350). New York: Springer.
- Okubo, F., Hirokawa, S., Terai, M., Shimada, A., Kentaro, K., Yamada, M., & Ogata, H. (2016, January). Learning activity features of high performance students. *CEUR Workshop Proceedings* (Vol. 1601, pp. 28-33).
- Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33.
- Romainville, M. (1994). Awareness of cognitive strategies: The relationship between university students' metacognition and their performance. *Studies in Higher Education*, 19(3), 359-366.
- Shimada, A., Okubo, F., Yin, C., Kojima, K., Yamada, M., & Ogata, H. (2015, July). Informal learning behavior analysis using action logs and slide features in e-textbooks. *2015 IEEE 15th International Conference on Advanced Learning Technologies* (pp. 116-117). IEEE.
- Tobias, S., & Everson, H. T. (2002). Knowing What You Know and What You Don't: Further Research on Metacognitive Knowledge Monitoring. *Research Report No. 2002-3. College Entrance Examination Board*.
- Tobias, S., & Everson, H. T. (2009). The importance of knowing what you know: A knowledge monitoring framework for studying metacognition in education. In *Handbook of metacognition in education* (pp. 119-140). New York: Routledge.
- Wang, T. H. (2011). Developing web-based assessment strategies for facilitating junior high school students to perform self-regulated learning in an e-Learning environment. *Computers & Education*, 57(2), 1801-1812.
- Van Horne, S., Russell, J. E., & Schuh, K. L. (2016). The adoption of mark-up tools in an interactive e-textbook reader. *Educational Technology Research and Development*, 64(3), 407-433.
- Yamada, M., Shimada, A., Okubo, F., Oi, M., Kojima, K., & Ogata, H. (2017). Learning analytics of the relationships among self-regulated learning, learning behaviors, and learning performance. *Research and Practice in Technology Enhanced Learning*, 12(1), 13.
- Yamada, M., Shimada, A., Oi, M., Taniguchi, Y., & Konomi, S. (2018). BR-MAP: Concept Map System using e-Book Logs. In 15th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2018), pp. 248-254.
- Yin, C., Okubo, F., Shimada, A., Oi, M., Hirokawa, S., Yamada, M., Kojima, K., & Ogata, H. (2015). Analyzing the features of learning behaviors of students using e-books. In Proceedings of the International Conference on Computers in Education (pp. 617-626).

- Yin, C., Yamada, M., Oi, M., Shimada, A., Okubo, F., Kojima, K., Ogata, H. (2019). Exploring the Relationships between Reading Behavior Patterns and Learning Outcomes based on Log Data from e-books: A Human Factor Approach, *International Journal of Human-Computer Interaction*, 313-322.
- Zimmerman, B. J., & Schunk, D. H. (Eds.). (2001). *Self-regulated learning and academic achievement: Theoretical perspectives*. New York: Routledge.

Recommendation of Personalized Learning Materials based on Learning History and Campus Life Sensing

Keita Nakayama^{1*}, Atsushi Shimada^{2*}, Tsubasa Minematsu^{3*},
Masanori Yamada^{4*}, Rin-ichiro Taniguchi^{5*}

*Kyushu University, Japan

¹nakayama@limu.ait.kyushu-u.ac.jp, ²atsushi@ait.kyushu-u.ac.jp,

³minematsu@limu.ait.kyushu-u.ac.jp, ⁴mark@mark-lab.net, ⁵rin@kyudai.jp

ABSTRACT: Thanks to the widespread of ICT environments not only in social life but also in the educational field, personalized learning has become a real possibility in recent years; it is expected to provide adaptive support to learners based on their situations. A typical approach is to recommend personalized learning materials based on the learning progress or understanding level of learners. Effective recommendations will maintain learners' motivation and help them overcome weaknesses. However, too many and/or too frequent recommendations sometimes interfere with learners' interests. Therefore, it is important to consider the timing of when recommendation information should be sent to learners. In other words, we assume that timely recommendations will encourage learners' motivation more than greedy strategies. In our study, we proposed a new strategy to support personalized learning. Our approach collaborates with activity sensing during campus life and automatically detects the timing of recommendations. Moreover, our recommendations provide a short summary of learning materials, which enhances learners' previews compared with the original materials. In this paper, we introduce the configuration of the proposed system, followed by a report of preliminary experimental results and a mention of future works.

Keywords: recommendation, adaptive learning, learning history, activity sensing

1 INTRODUCTION

With the development of information technology, various ICTs have been introduced into the educational and learning environment, along with increasing expectations for the realization of personalized learning environments that can recommend appropriate learning materials (LMs) based on the learner's history of learning and provide support to enhance the learning effect (Hwang et al, 2017; Yin & Hwang, 2018). On the other hand, the teaching style is still the traditional face-to-face and multi-person simultaneous lecture style. For this reason, it is difficult to conduct lectures based on the learning progress and understanding level of individual learners. Furthermore, with the diversification of learning methods and lifestyles, providing the same LMs to all learners is not appropriate. Therefore, it is necessary to realize adaptive learning support that matches individual situations (Truong, 2016).

One approach to learning support based on the individual learning situation is the personalized recommendation of digital LM based on the learning situation. Many traditional methods recommend the use of LMs in a lecture or self-learning (e.g., Lan & Baraniuk, 2016; Wan & Niu, 2018). Thus, it is assumed that the learner is learning at the time of recommendation. To provide further learning support, it is important to use not only conventional methods but also to promote learning in daily life even outside of learning time. To encourage learning outside of learning time, it is important for the

system to consider the timing to recommend (Pielot et al, 2015). With the spread of smartphones, it has become possible to make use of data measuring daily life and these data has become using in recommendation method (Okoshi et al., 2019). Using these technologies, it will be possible to promote learning appropriately in daily life outside of study hours.

This study aims to provide learning support that effectively occupies learners' free time by individually recommending LMs based on learning history and campus life sensing, using university education as a test environment. In this study, we propose a learning support system that recommending lecture slides as digital LMs composed of suitable contents and quantities for each learner at appropriate times by integrating the three methods of detection by a smartphone of free time when the learner may study, recommendation based on learning history and learner's activity and automatic summarization of digital LMs. In the following section of this paper, we introduce the configuration of the proposed system and each method followed by a report of preliminary experimental results.

2 PERSONALIZED RECOMMENDATION SYSTEM BASED ON LEARNING HISTORY AND CAMPUS LIFE SENSING

2.1 System overview

In this study, we propose a learning support system that recommends lecture slides as digital LMs composed of suitable content and amount for each learner at an appropriate time based on the learner's current activity and learning history. The proposed system takes three steps. Step 1 is that learning time is detected by a sensor in a learner's smartphone. Step 2 is that LM is determined based on the learner's learning history. Step 3 is that the LM determined in the step 2 is recommended to the learner via email at the time detected by the step 1.

Figure 1 shows the configuration of the proposed system. Moodle (Flanagan & Ogata, 2018) is a learning management system. BookRoll (Flanagan & Ogata, 2018) is an e-Book system which has various data such as LMs, learning logs, etc. Teachers and students access BookRoll via Moodle and browse LMs. This system detects the learner's free time to study and recommends LMs based on the learner's learning history. The systems for realizing this function are an application measuring the learner's activity and a campus activity server (CAS). The application is installed on the learner's smartphone and measures the learner's activity using the sensor installed in the device. Based on the measured data, the application detects whether the learner is ready to learn and notifies the CAS. The CAS manages recommendations to learners. The CAS determines the recommended LM and timing from the learning logs recorded in BookRoll and the information notified by the application and then sends an email. This is to encourage learning at times when the learner does not intend to study, which can be a significant burden on the learner. It is therefore desirable that the recommended LMs enable the learning of important contents in a short period of time. For this reason, this study recommends LMs with summarized contents. Shimada et al (2016) proposed automatic summarization method based on the advance organizer theory (Ausubel, 1960), and suggested that short summaries may have enhanced students' motivation to preview the material. Therefore, it is considered that the summarized version of LM has less burden on the learner. Our summarization system is inspired by (Shimada et al., 2016) and can summarize LMs registered in BookRoll. Details of the application measuring learner's activity, CAS, and summarization system are described in the following sections.

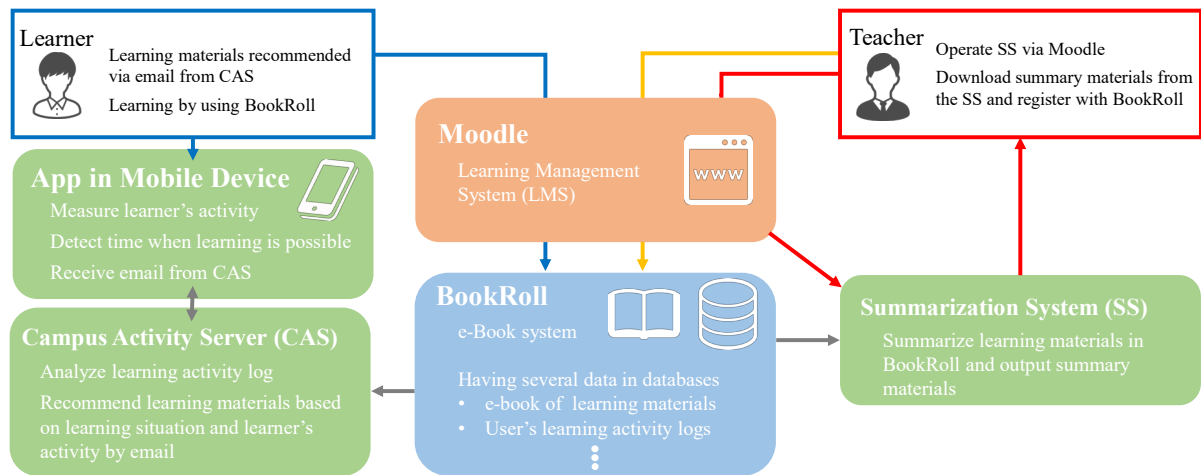


Figure 1: System configuration

2.2 Learning time detection

In this study, we recommend LMs via email and encourage learning during the learner's free time. The reason for measuring learner's activity is that the effect of recommendation timing on the results has been clarified in research such as ubiquitous computing (Pielot et al, 2015), and the recommendation effect is higher when the user's situation is taken into account (Okoshi et al., 2019). We have developed an application to measure learner activity and detect the free time when the learner may study. Most learners usually always carry smartphones, and the movement of smartphones is thought to be influenced by the learner's activity. Therefore, the activity of the learner is measured using the sensor installed in the smartphone. In this study, we use an Android phone and measure the learner's activity using the acceleration sensor. The Android phone can measure acceleration along three axes. In this study, it is assumed that the learner may study when the terminal is not moving so hard. This is because, it is considered that the timing when the learner may study (such as not moving, on the bus, etc.) is when the terminal is in a stable state (Li et al, 2013). When the measured three-axis acceleration does not exceed a certain threshold value for a certain period of time, the developed application detects that the learner is ready to learn and notifies the CAS.

2.3 Recommendation based on learning history and learner activity

The campus activity server (CAS) recommends LMs to each learner based on the learning history collected from the BookRoll learning logs and activity status obtained from the application. The recommendation method takes four steps. First, the CAS receives notification from the application that the learner may study. Second, the CAS checks the recommendation history logs and decides whether to recommend. Third, the CAS decides which LMs to recommend based on learning history logs stored in BookRoll. Finally, the CAS sends an email recommending LMs. This server manages recommendations based on two types of logs: One is a history of recommendation to learners, while the other is the learner's browsing history for recommended LMs. First, each log is explained, and then the recommendation method is explained.

The record of a recommendation to a learner is recorded on the server when an email is sent. The recorded information includes the ID of the recommended user, the ID of the recommended LM, the recommended time, and the number of pages of recommended LM. The logs of a learner's browsing

User ID	e-book ID	e-book title	Page	Operation	...	Last_read time
xxxxxxx	oooooooo	*****	5	NEXT	...	2019-09-01 11:26:15
xxxxxxx	oooooooo	*****	6	NEXT	...	2019-09-01 11:26:50
xxxxxxx	oooooooo	*****	7	PREV	...	2019-09-01 11:27:45
xxxxxxx	oooooooo	*****	6	NEXT	...	2019-09-01 11:28:35
xxxxxxx	oooooooo	*****	7	CLOSE	...	2019-09-01 11:30:15

Figure 2: Learning logs stored in BookRoll

User ID	e-book ID	Page	Browsing_time	Read_flg	Last_read_time
xxxxxxx	oooooooo	1	100	1	2019-09-01 **:**:**:
xxxxxxx	oooooooo	2	120	1	2019-09-01 **:**:**:
xxxxxxx	oooooooo	3	60	1	2019-09-01 **:**:**:
xxxxxxx	oooooooo	4	0	0	
xxxxxxx	oooooooo	5	0	0	

Figure 3: learner's browsing history logs

history are a collection of learning logs for the LMs stored in BookRoll. Figure 2 shows samples of learning logs in BookRoll. The learning logs record operations such as page transitions for digital LMs. The CAS counts the browsing time of each page of the recommended LM when deciding which LM to recommend. The log of a learner's browsing history is shown in Figure 3. In Figure 3, "Browsing time" indicates the number of seconds T of accumulated browsing time after recommendation, while "Read flag" indicates whether learning has been completed. When the contents of page i of LM has been learned, the value of "Read flag" f_i is 1; otherwise, it is 0.

Next, the recommendation method is described. First, the server is notified by the application. When the server receives the notification, it checks the recommendation history logs and decides whether to recommend. In this study, the current time is compared with the last recommended time, and the LM is recommended to the learner if it has not been recommended for a certain period of time. When recommending to the learner, the LM to be recommended is determined from the logs of the learner's browsing history. In this paper, as an initial stage of study, the server recommends LMs in a predetermined order. When learning with the LM is completed after the recommendation, the system recommends the next item in the order of LM. In this paper, when each page of the LM is viewed ($T_i > 0$), it is assumed that the content of the page is learned ($f_i = 1$). If all the pages of the recommended LM are not viewed, the LM recommended the last time is recommended again. Finally, the LM determined by the above method is recommended by email.

2.4 Summarization of learning materials

In this section, we provide an overview of the automatic summarization system (Shimada et al., 2016). This system was designed to produce a set of lecture slides. The purpose of slide summarization is to select a subset of pages that maximizes the importance of content under a given condition (in this case, browsing time). To achieve this, the system selects the important pages without losing the overall narrative of the lecture. In this section, we will give an overview of the summarization method.

First, lecture material is analyzed to extract important visual and textual features from each page. In terms of visual importance, the number of objects such as text, figures, formulas, etc. in each material is estimated, using a background subtraction technique and an inter-frame difference technique. Word importance is estimated using the TF-IDF method. Furthermore, a teacher specifies the desired browsing time for students to study each page. The visual, textual, and temporal features determined by these methods are then combined to generate an importance score. Finally, an optimal subset of pages is selected, which maximizes the importance score for learning in a short time.

3 EXPERIMENT

We conducted preliminary experiments to confirm the performance of the proposed system and students' reactions to recommendations. The preliminary experiment was conducted beforehand in

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

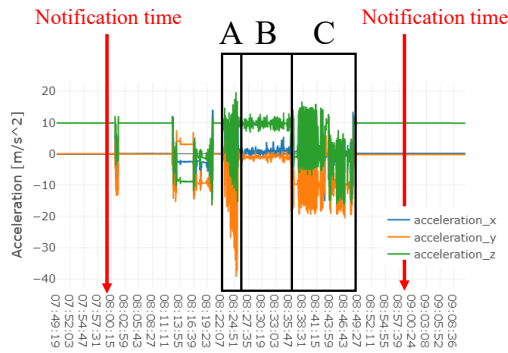


Figure 4: Sample of sensing data

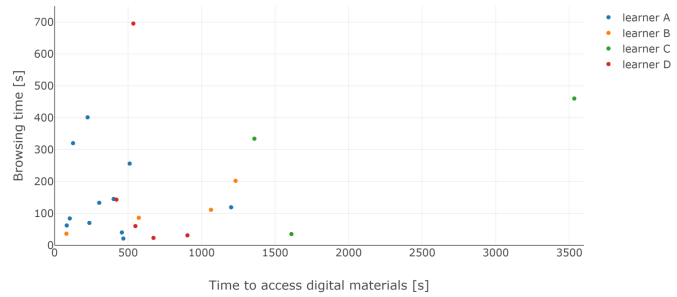


Figure 5: The distribution of the access timing vs. the browsing period

about 10 hours for 4 university students on campus. In this experiment, if the acceleration of the three axes did not change more than 5 m/s^2 for about 10 minutes, the CAS was notified that it was free time for learning with the application. Based on the notification, an email was sent recommending LMs with an interval of at least one hour. The email contains the title of the recommended LM and a link to Moodle. The recommended LMs are summaries of the materials that will be used in the data science course in the second semester of 2019/2020 at our university. The recommended LMs consisted of 8-16 pages, 20% of the original LMs. In this experiment, the response to the recommended email was arbitrary considering the actual usage situation.

Figure 4 shows a sample of acceleration change of a subject. The change in acceleration in the Figure 4 shows that part "A" is running, part "B" is on the bus, and part "C" is walking. From this result, it was confirmed that various human activities can be observed by the application. From this measurement result, the application notified the server that the learner's free time was between 8 and 9 a.m. Note that with the proposed method, LMs may be recommended if the status of Part B persists for a certain period of time. In this way, it was confirmed that the developed application could recommend LMs at a time that seemed to be the learner's free time.

As a result of the experiment, the LMs were recommended 35 times for the learners, and the recommended LMs were accessed 23 times. The average time from receiving the email to accessing the recommended digital LMs was 12 minutes 4 seconds. From this result, it can be seen that the LMs were accessed in a relatively short time after receiving the email, and the recommendation was conducted when the learner was able to study. The average browsing time for each recommended LM was 3 minutes 41 seconds. From this result, we can see that each LM can be browsed in a short time as we intended. Figure 5 visualizes the scatter plots of the access timing vs. the browsing time. Note that the same LM was recommended multiple times, so that the number of recommendation was higher than the number of LMs prepared in advance. The x-axis represents the time from when the email was sent to the learner until the digital LM was accessed, and the y-axis is browsing time per access to the LM. From Figure 5, we can see there is no correlation between the access timing and browsing time. This is an expected result because the proposed method sent the recommendation information when the subject was estimated to have time, and whether he/she accessed to the LM is completely up to the subject. From the results that the subjects accessed the materials less than 600 seconds in most cases, the timing of recommendation was almost suitable to the learners, and the learner studied using the free time. In the future, we plan to experiment with more learners and verify the educational usefulness of the proposed system.

4 CONCLUSION

In this paper, we proposed a learning support system by recommending learning materials based on learning history and campus life sensing. The proposed system recommends learning materials suitable for the learner based on the learning history during the learner's free time detected by the smartphone. In the previous experiment, a small number of subjects used the system, and we confirmed the behavior of the system and the response to recommendations from learners.

As future work, we plan to conduct long-term experiments on more subjects using the proposed system, based on the results of this preliminary experiment. In this paper, as an initial stage of study, learning materials were recommended in a predetermined order based on learning history. We will also consider how to recommend learning materials using other types of data such as student knowledge information, quiz results recorded in Moodle, and so on.

ACKNOWLEDGEMENTS

This work was supported by JST AIP Grant Number JPMJCR19U1, iLDi Grand Challenge 2018-#1, and JSPS KAKENHI Grand Number JP18H04125, Japan.

REFERENCES

- Ausubel, D. P. (1960). The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of educational psychology*, 51(5), 267.
- Flanagan, B., & Ogata, H. (2018). Learning analytics platform in higher education in Japan. *Knowledge Management & E-Learning: An International Journal*, 10(4), 469-484.
- Hwang, G. J., Chu, H. C., & Yin, C. (2017). Objectives, methodologies and research issues of learning analytics. *Interactive Learning Environments*, 25(2), 143-146.
- Lan, A. S., & Baraniuk, R. G. (2016, June). A Contextual Bandits Framework for Personalized Learning Action Selection. In EDM (pp. 424-429).
- Li, M., Ogata, H., Hou, B., Uosaki, N., & Mouri, K. (2013). Context-aware and Personalization Method in Ubiquitous Learning Log System. *Journal of Educational Technology & Society*, 16(3).
- Okoshi, T., Tsubouchi, K., & Tokuda, H. (2019, July). Real-World Product Deployment of Adaptive Push Notification Scheduling on Smartphones. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2792-2800). ACM.
- Pielot, M., Dingler, T., Pedro, J. S., & Oliver, N. (2015, September). When attention is not scarce- detecting boredom from mobile phone usage. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing* (pp. 825-836). ACM.
- Shimada, A., Okubo, F., Yin, C., & Ogata, H. (2017). Automatic Summarization of Lecture Slides for Enhanced Student Preview Technical Report and User Study. *IEEE Transactions on Learning Technologies*, 11(2), 165-178.
- Truong, H. M. (2016). Integrating learning styles and adaptive e-learning system: Current developments, problems and opportunities. *Computers in human behavior*, 55, 1185-1193.
- Wan, S., & Niu, Z. (2018). An e-learning recommendation approach based on the self-organization of learning resource. *Knowledge-Based Systems*, 160, 71-87.
- Yin, C., & Hwang, G. J. (2018). Roles and strategies of learning analytics in the e-publication era. *Knowledge Management & E-Learning: An International Journal*, 10(4), 455-468.

OpenLA: An Open-Source Library for e-Book Log Analytics

Ryusuke Murata, Tsubasa Minematsu, Atsushi Shimada
Kyushu University
{murata, minematsu, atsushi}@limu.ait.kyushu-u.ac.jp

ABSTRACT: This paper proposes a new open-source library for learning analytics, named “OpenLA”. The OpenLA provides five modules; Basic Module, Data Extraction Module, Data Aggregation Module, Data Conversion Module, and Data Visualization Module. Each module has several fundamental APIs which support a development of learning analytics coding. The OpenLA assumes to collaborate with the dataset format, which is based on e-Book event logs, given in LAK Data Challenge Workshop in 2019 and 2020. We believe that the APIs will assist and accelerate researches of analytics.

Keywords: OpenLA, open-source library, learning analytics, e-Book event logs

1 INTRODUCTION

Thanks to the widespread of ICT environment and digital learning system, it became possible to collect not only learning results such as examination results, but also learning process how each learner studies a content, how long he/she spent time for study, etc. Understanding the behaviors of learners is a crucial issue in the learning analytics (LA), so that learning logs collected via digital learning systems are often utilized for analytics of teaching and learning. An e-Book system is one of the useful systems which records learning process such as when a learner opens a learning material, or turns a page in the material. The operation logs are primitive actions on the learning material, and preprocessing is necessary to convert the original logs into more sophisticated representations for finding good features in learning activities (Yin 2019), understanding the behavior of learners (Shimada 2019), estimating academic performance (Okubo 2018), discovering at-risk students (Shimada 2018), and so on. So far, such preprocessing was developed by each researcher even though there are many common processes; calculating reading time of each learner, counting up events operated by a specific learner, page-wise summary of event operations, etc. To reduce the redundant development of these preprocessing functions and accelerate the development of core technologies for advanced learning analytics, we have developed an open source library, so called “OpenLA”. The OpenLA provides essential APIs to analyze e-Book event logs collected by BookRoll(Ogata 2015); data conversion, data extraction, data aggregation, and data visualization. In this paper, we introduce the primal version of OpenLA.

2 API CONCEPT

OpenLA developed to analyze event logs which are open to public to conduct data challenge workshops in LAK19 and LAK20. The dataset includes four types of files:

Course_#_EventStream.csv

- Data of the logged activity data from learner' interactions with the BookRoll system.

Course_#_LectureMaterial.csv

- Information about the length of the lecture materials used.

Course_#_LectureTime.csv

- Information about the schedule of the lectures.

Course_#_QuizScore.csv

- Data on the final score for each student.

The APIs are written in Python language. The supported version is Python 3.7.X and required libraries are pandas 0.25.X, numpy 1.16.X and matplotlib 3.1.X.

The meanings of arguments used in each function are as follows:

course_id: indicates a specific course.

contents_id: indicates a specific content, i.e., a lecture material.

lecture_week: indicates a specific lecture, i.e., what week the lecture was conducted.

user_id: indicates a specific learner

2.1 Basic Module

The Basic Module loads the dataset and gets basic information about courses, lecture time, users. Eventstream data is loaded as pandas.DataFrame format.

```
def load_eventstream(self):
```

```
:return: DataFrame containing EventStream (pandas.DataFrame).
```

```
def contents_ids_in_course(self):
```

```
:return: list of contents ids (list of str).
```

```
def contents_id_to_lecture_week(self, contents_id):
```

```
:return: lecture week(s) of the contents_id(s) (int or list of int).
```

```
def lecture_week_to_contents_id(self, lecture_week):
```

```
:return: contents id(s) of the lecture week(s) (str or list of str).
```

There are the other functions which have the same design with above functions. Due to the page limitation, we just list the name of the other functions. For the details, refer to our website (<https://www.leds.ait.kyushu-u.ac.jp>).

```
def contents_id_to_num_pages(self, contents_id):
```

```
def lecture_week_to_num_pages(self, lecture_id):
```

```
def contents_id_to_start_time(self, contents_id):
```

```
def lecture_week_to_start_time(self, lecture_week):
```

```
def contents_id_to_end_time(self, contents_id):
```

```
def lecture_week_to_end_time(self, lecture_week):
```

```

def user_num_in_course(self):
    :return: number of users in course(int).

def users_in_course(self):
    : return: list of user ids in course (list of str).

def user_score(self, user_id):
    : return: quiz score(s) of the user(s) (float or list of float). If “user_id” is NULL, the function
    returns all learners’ scores.

def users_in_selected_score(self, users_list=None, bottom=None, top=None)::
    : return: list of user ids whose scores are in the range between bottom and top (list of str).

```

2.2 Data Extraction Module

The Data Extraction Module receives DataFrame of EventStream, and extracts information required.

```

def select_user(self, dataframe, user_id):
    :return: Extracted result (pandas.DataFrame). If the argument “user_id” is given as a list of user
    ids, the function extracts all the users in the list.

```

The other functions in Data Extraction Module are also designed in the same manner. We just introduce the name of functions due to the page limitation, but refer to our website (<https://www.leds.ait.kyushu-u.ac.jp>) for the details.

```

def select_contents(self, dataframe, contents_id):
def select_operation(self, dataframe, operation_name):
def select_marker_type(self, dataframe, marker_type):
def select_device(self, dataframe, device_name):
def select_page(self, dataframe, bottom=None, top=None):
def select_memo_length(self, dataframe, bottom=None, top=None):
def select_time(self, dataframe, start_time, end_time):
def select_by_lecture_time(self, dataframe, contents_id=None, timing='during'):

```

2.3 Data Aggregation Module

The Data Aggregation Module receives DataFrame, and returns aggregation results.

```

def user_num_in_dataframe(self, dataframe):
    :return: number of user ids in dataframe (int)

def users_in_dataframe(self, dataframe)
    :return: list of user ids in dataframe (list of str)

```

The other functions listed below receive the same arguments, and return aggregation results in “list of str” format.

```
def contents_ids_in_dataframe(self, dataframe):
def operations_in_dataframe(self, dataframe):
def marker_types_in_dataframe(self, dataframe):
def device_codes_in_dataframe(self, dataframe):
```

2.4 Data Conversion Module

The Data Conversion Module receives DataFrame of EventStream, and calculate the number of events, page transition, reading time of each page, etc. The calculation result is also acquired by DataFrame format.

```
def users_operation_count(self, dataframe):
    :return: "user id" vs. "operation"s, i.e., how many times each learner used each operation
    (pandas.DataFrame, columns['user id', 'each event']).

def users_page_transition_and_event_count(self, dataframe):
    :return: "user id" vs. "contents id", "page no", "operation"s with consideration of page transition,
    i.e., how many times each learner used each operation in each page and in each content. The
    counting is performed every page transition (pandas.DataFrame, columns['user id', 'contents id',
    'page no', 'each event']).

def users_pageno_and_event_count(self, dataframe):
    :return: "user id" vs. "contents id", "page no", "staying time", i.e., how many times each learner
    used each operation in each page and in each content. The result is equivalent to the page-wise
    aggregation of "users_page_transition_and_event_count" function (pandas.DataFrame,
    columns['user id', 'contents id', 'page no', 'each event']).

def users_page_transition_and_staying_seconds(self, dataframe):
    :return: "user id" vs. "contents id", "page no", "operation"s with consideration of page transition,
    i.e., how long each learner stayed in each page and in each content. The calculation is performed
    every page transition (pandas.DataFrame, columns['user id', 'contents id', 'page no',
    'time_to_enter', 'time_to_leave', 'staying_seconds']).

def users_pageno_and_staying_seconds(self, dataframe):
    :return: "user id" vs. "contents id", "page no", "operation"s, i.e., how long each learner stayed in
    each page and in each content. The result is equivalent to the page-wise aggregation of
    "users_page_transition_and_staying_seconds" function (pandas.DataFrame, columns['user id',
    'contents id', 'page no', 'staying_seconds']).
```

2.5 Data Visualization Module

The Data Visualization Module receives DataFrame, and make a visual graph.

```
def scatter_graph(dataframe, column_x, column_y, save_file=None):
    :return: Draw scatter plots between two columns in DataFrame. If the "save_file" is indicated,
    the graph is saved.
```



```
def time_series_graph(dataframe, column, graph_type='line', start_time=None, end_time=None,
    save_file=None):
    :return: Draw a line series graph of indicated "column". If the "save_file" is indicated, the graph
    is saved.
```

```
def operation_count_bar_graph_in_pages(self, user_id, contents_id, operation_name,
    save_file=None):
    :return: Draw a bar graph which represents page-wise counting result of each operation used by
    a specific learner. If the "save_file" is indicated, the graph is saved.
```

3 USAGE EXAMPLE

We show an example how to use the OpenLA APIs. The code in Figure 1 to investigate a student who got a good quiz score, and how he/she browsed the lecture material during the lecture time.

After executing the above code, a line graph is drawn as shown in Figure 2.

```
from OpenLA import OpenLA

# Load an EventStream data of a specific course
c1 = OpenLA(files_dir="dclak20/Train", course_id="6b1900c56c")
event_stream_df = c1.load_eventstream()
user_id_list = c1.users_in_course()
contents_id_list = c1.contents_ids_in_course()
score_list = c1.user_score()

# Get the user id who got the best score
top_score = max(score_list)
top_score_users = c1.users_in_selected_score(users_list=user_id_list, bottom=top_score,
    top=top_score)
a_top_score_user = top_score_users[0]

# Extract the event logs of the learner
top_score_user_df = c1.select_user(dataframe=event_stream_df,
    user_id=a_top_score_user)

# Extract the operations of a specific lecture material (contents_id) during the lecture time
top_score_in_lecture_df = c1.select_by_lecture_time(dataframe=top_score_user_df,
    contents_id=contents_id_list[0], timing='during')

# Draw a line graph
c1.time_series_graph(dataframe=top_score_in_lecture_df, column='pageno',
    graph_type='line', start_time=None, end_time=None, save_file='example.png')
```

Figure 1: Sample code using OpenLA

4 CONCLUSION

This paper proposed a new open-source library "OpenLA" for analytics of e-Book logs. The APIs are useful to load the dataset coming from BookRoll system, to extract learner-wise, content-wise, lecture-wise information, to aggregate and convert the original event logs to a given format. We

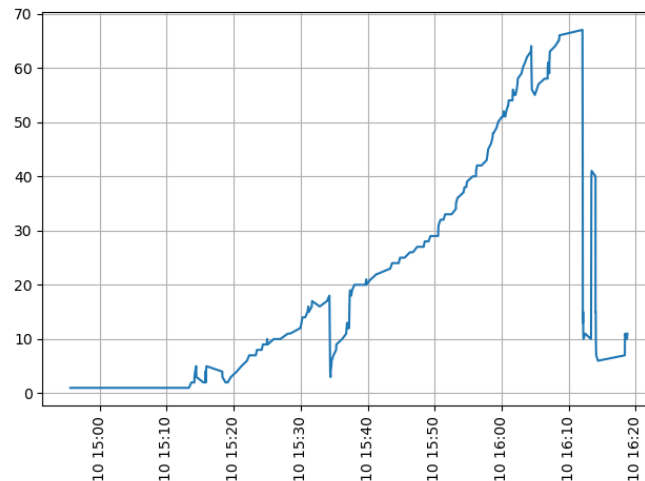


Figure 2: A line graph drawn by OpenLA

believe that the OpenLA APIs support and accelerate the development of learning analytics methodologies based on e-Book log analytics. The latest information will be post on our website: <https://www.leds.ait.kyushu-u.ac.jp>.

ACKNOWLEDGEMENTS

This work was supported by JST AIP Grant Number JPMJCR19U1, and JSPS KAKENHI Grand Number JP18H04125, Japan.

REFERENCES

- Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). *E-Book-based learning analytics in university education*. In International Conference on Computer in Education (ICCE 2015), pp. 401-406.
- Okubo, F., Yamashita, T., Shimada, A., Taniguchi, Y., Konomi, S. (2018). *On the Prediction of Students' Quiz Score by Recurrent Neural Network*. CrossMMLA@LAK 2018
- Shimada, A., Taniguchi, Y., Okubo, F., Konomi, S., Ogata, H. (2018). *Online change detection for monitoring individual student behavior via clickstream data on E-book system*. International Conference on Learning Analytics and Knowledge (LAK18).
- Shimada, A., Mouri, K., Taniguchi, Y., Ogata, H., Taniguchi, R., Konomi, S. (2019). *Optimizing Assignment of Students to Courses based on Learning Activity Analytics*. International Conference on Educational Data Mining (EDM2019).
- Yin, C., Yamada, M., Oi, M., Shimada, A., Okubo, F., Kojima, K., Ogata, H. (2019). *Exploring the Relationships between Reading Behaviour Patterns and Learning Outcomes Based on Log Data from E-Books: A Human Factor Approach*, International Journal of Human-Computer Interaction, pp. 313-322.

Learning Engagement - Clustering Analysis based on Student Interaction with Digital Textbooks.

Abu Abu Eyo¹, Owoeye Oluwaseyi², Ocheja Patrick³, Flanagan Brendan², and Ogata Hiroaki²

¹ University of Aberdeen, UK

² WealthTech Limited, Nigeria

³ Kyoto University, Japan
ocheja.ileanwa.65s@st.kyoto-u.ac.jp

ABSTRACT: The logging of student reading activities, facilitated by the increasing use of digital textbooks has made it possible to track the level of engagement students have with course material and identify patterns from this data. In this paper several features were extracted from data logged on the Bookroll application and used as metrics in carrying out analysis. Clustering of students was carried out to identify subgroups of students based on their scores and how their engagement with the application reflected on their performance, prediction of students' quiz scores based on their interaction with the application was also carried out. The significance of this work lies in its potential to help academics with insight on the engagement traits of different subgroups of students on a course as well as pre-empt the performance of individual student on a course. This insight will help in enabling instructors spot problems early in the term and intervene by making alternative arrangements to ensure the success of different students on a course.

Keywords: learning engagement, students' performance, clustering, pre-class, post-class, ebook logs, learning logs, digital textbooks

1 INTRODUCTION

Learning analytics deals with the measurement, collection, analysis and reporting of data about learners to enable the understanding and optimization of learning and the environment and channels through which this occurs (Siemens, 2010). Projects in learning analytics rely on data collection from student engagement with the course via different channels such as class participation and engagement with course material to predict student performance outcomes, the latter being the focus of this work. Variables representing student reading engagement with course material have proved potent in carrying out analysis on student engagement with courses. In this paper several variables were extracted from a digital textbook application known as Bookroll and used as metrics in the measurement of reading engagement in a digital notebook environment. Two types of analysis were carried out based on these metrics and the final score awarded to each student on the course.

1.1 Digital Textbook

The use of digital textbooks by youths has been on the rise in recent years (Association of American Publishers, 2014) and this has led to the adoption of digital textbooks in higher education (Longhurst, 2003) despite conflicting findings in terms of the preferences of students. The growing use of digital textbook is providing academics with an unobtrusive means of collecting data about students' use of

educational material. Although the functionality of logging reader activities in digital textbook environments have been available for a long time, there has been a recent rising interest in the development of methods for the analysis of this data to help in predicting student performance outcomes. A recent digital textbook reading technology called BookRoll, developed at Kyoto University offers readers different interaction possibilities such as highlighting, note adding, marking of material as important or difficult, etc., enabling student reading behaviours to be tracked and logged for further analysis (Flanagan & Ogata, 2017; Ogata et al., 2015).

1.2 Engagement with Digital Textbook and Students' Performance

The relationship between student engagement with textbooks and their performance on the course was investigated by Astin (1984). Astin's (1984) thesis was that student engagement involves psychological energy investment by students and that the amount of learning achieved is proportional to the level of engagement in an educational activity. Also, Astin indicated that tests are a measure of the level of learning attained: students who invest more psychological energy in reading learn more and hence perform better at tests. Going by this, students more engaged with a course through reading should perform better and additional engagement made possible with digital textbooks such as highlighting, addition of memos and marking should lead to improved outcomes. Thus, this further strengthens the use of analysis on digital textbooks to directly measure the level of student engagement and predict student performance outcomes (Bossaller & Kammer, 2014).

1.3 Research Questions

The current study relies on logs generated from the BookRoll application to answer two key research questions:

1. What are the digital textbook reading patterns of students?
2. How does the performance of students on courses vary with their engagement with the digital textbook system?
3. Can we predict the performance of students on a course based on derived features from student interaction with the digital textbook system and how can these be used to help students perform better.

2 RELATED WORK

There have been a lot of recent work carried out in learning analytics. Various analysis with different aims have been pursued. This section provides an overview of some of the notable work, their aims and methods.

Whereas, methods of data collection have been similar, different approaches to analysis have been applied due to differences in aims. Atsushi et. al (2016) used a predefined number of patterns to explain the browsing patterns of students. The result of this work indicated that there is a relationship between browsing patterns and the level of understanding of contents. Atsushi et. al. (2019) used clustering to optimize the assignment of students to different courses to reduce the mismatch between students learning behavior and patterns and teachers. The novelty of the work lies in the demonstration of student score improvement after optimization of student assignment to courses. The overarching aims of these studies have been to identify patterns in learning behaviours of

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

students, the courses they take as well as their performance on those courses, to ensure that students with similar learning behaviours are assigned to the right courses and lecturers, increasing their chances of success at those courses.

In this work, the focus is on identifying subpopulations of students on a course as well as predicting their performance on the course. These two analyses when applied early on a course will enable educators identify problem areas and improve the chances of success of students on a course.

3 METHOD

3.1 Data Collection and Preprocessing

The data used for analysis in this paper was obtained from the Data Challenge workshop at the 10th Learning Analytics and Knowledge Conference. The data included anonymized click-stream data from 1326 students distributed across 10 different courses whose learning contents are accessed through BookRoll. BookRoll system records students' engagement activities when they use a learning content. The provided data also contained information on the learning contents used for each course, the lecture schedule, students' engagements and quiz scores. In this paper, we carried out a subpopulation analysis of students in the provided dataset. To answer the identified research questions, we outlined the following analysis:

- a) Identify students' reading activities towards predicting their performance.
- b) Identify the different student subpopulations based on multiple features including sessions, interactions, page activities, etc. It is expected that findings here will corroborate initial findings from other authours.
- c) Predict student performance based on features derived from logs of student interaction with the digital textbook.

3.2 Learning Activity Summary

To investigate students' activities, we analyzed the interactions students had with course material by getting statistical summaries of actions taken on the learning contents provided for each course. The aggregation procedure was done on a course by course basis. We derived features from the raw data set such as *MEM_LEN*, *SEARCH*, *ADD_BMRK* and *SCORE* as described in table 1.

Table 1: Description of digital textbook reading variables for analysis.

Variable	Description
MEM_LEN	Total length of memos added by the student.
SEARCH	Total number of times a student searched the learning content.
ADD_BMRK	Total number of bookmarks added by the student.
SCORE	Score obtained by the student on a given course.

3.3 Clustering Analysis

In this paper, k-means clustering (MacQueen 1967) was used to identify the different groups in the derived datasets. We combined the data from the ten different courses into one by applying aggregate operations to obtain new variables for clustering analysis. We performed this initial aggregation of the provided dataset to get a better understanding of student behavior across board and to identify the different subpopulations within our dataset. This was done by computing aggregates of the various attributes per student per course to obtain the variables described in table 2. A total of fourteen features were obtained from the raw data and the elbow method (Bholowalia & Kumar, 2014) was used to determine the optimal number of clusters of 5 as shown in figure 1. The elbow method is based on the percentage of variance where a given number of clusters is adjudged optimal if adding an additional cluster does not result in a better modelling. Initial plot of obtained clusters showed 74 outliers which were detected and removed using the z-score (normalization) of each point with respect to their cluster's centroid.

Table 2: Description of digital textbook reading derived variables from aggregates (N = 1253).

Variable	Description	Mean (Standard Deviation – SD)
SCORE	Score obtained by the student in a given course.	83.59 (7.77)
SESSION	Total number of reading sessions.	16.38 (7.51)
INTERACTIONS	Total number of interactions a student had with the application.	1430.06 (853.25)
SEARCH	Total number of times a student searched the learning content.	1454.25 (846.22)
ADD_MRK	Total count of markers added to the learning material.	64.81 (105.49)
ADD_MRK_I	Total count of important markers added.	63.08 (107.97)
ADD_MRK_D	Total count of difficult markers added.	10.86 (30.16)
MEM_LEN	Total length of memos added by the student.	591.30 (1506.72)
ADD_BMRK	Total number of bookmarks added by the student.	6.34 (15.42)
ADD_MEM	Total count of memos added by the student.	6.47 (16.22)
DEL_MRK	Total count of deleted markers by student	9.38 (16.64)
PAGE_JMP	Total number of page jumps made by student.	33.54 (32.69)
PREV_PAGE	Total number of backward page movement by student.	423.40 (319.03)
NEXT_PAGE	Total number of forward page movements by student.	851.29 (466.67)

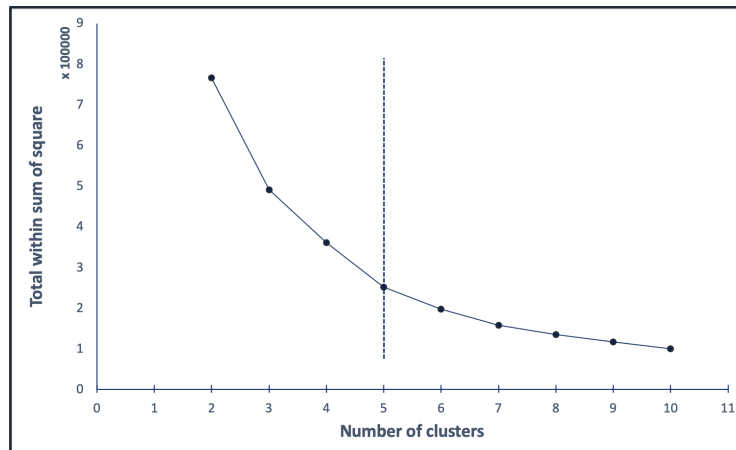


Figure 1: Optimal number of clusters detection using elbow method

3.4 Student Performance Prediction

The third phase involved learning from the data and building a regression model to predict students' performance based on the features derived from their interactions with the application. The same features used in the clustering were used in carrying out regression analysis to predict individual students' scores.

4 RESULTS AND DISCUSSIONS

4.1 Results of clustering analysis

In this section, we present the results obtained from the analysis concluded so far. The results obtained from the clustering analysis is shown in table 3. 14 features were used for the clustering but only 6 have been reported in table 3. for brevity. A plot of the 5 clusters is shown in figure 2. The T-Distributed Stochastic Neighbour Embedding (T-SNE) was used in visualizing the dataset. The T-SNE plot could be computationally expensive when dealing with datasets with high dimensionality, however, since we used just 14 features in our analysis, we leveraged T-SNE which looked for how best to represent the data using an optimal number of features. We now discuss the insights drawn from each of the 5 clusters obtained.

Table 3: Clusters of students based on digital textbook events and performance (N = 1253).

CLUSTER	N	SCORE	SESSION	SEARCH	ADD_MRK	ADD_MRK_I	MEM_LEN
1	216	85.52	17.71	1735.72	135.98	131.38	2876.04
		(6.84)	(8.09)	(837.66)	(143.15)	(143.35)	(2574.33)
2	339	83.03	17.14	1463.09	60.54	58.57	99.52
		(7.47)	(5.54)	(217.49)	(95.19)	(100.17)	(166.96)

		82.72	9.66	439.59	7.60	6.67	8.2
3	171	(9.91)	(4.83)	(154.20)	(26.11)	(26.66)	(31.74)
		82.64	13.26	870.52	18.46	17.67	90.26
4	284	(8.15)	(3.74)	(137.27)	(37.71)	(39.76)	(149.11)
		84.40	22.54	2587.95	101.90	101.42	242.40
5	243	(6.38)	(8.61)	(698.42)	(117.49)	(123.94)	(369.45)

Cluster 1: This group consists of 216 students with the highest mean score and second highest mean session time. Students in this cluster showed a likely trend of high engagement leading to high performance and students in the cluster can be classified as High Engagement High – Performing students. Across the 5 clusters, these students added the greatest number of ADD_MRK_I event which means that they identified important concepts more and highlighted such information than any other cluster. Also, this group of students have more lengthy memos (MEM_LEN) than students in other clusters. This finding is also indicative of high-performance and further tests can be carried out to determine the significance of this correlation.

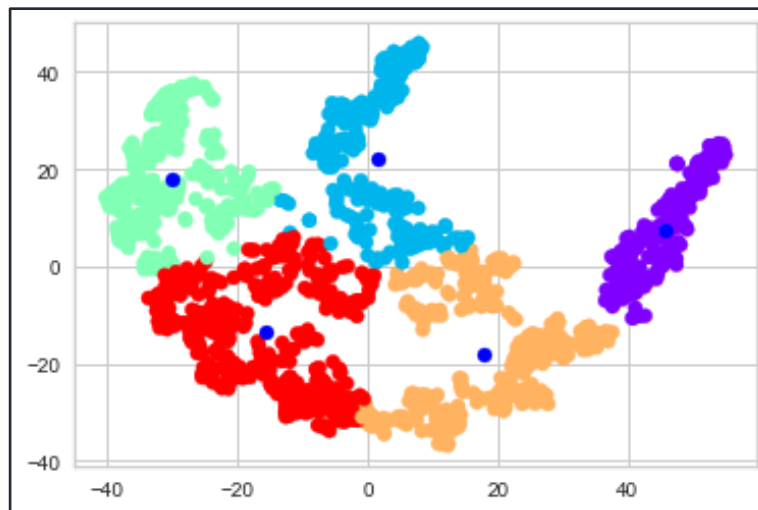


Figure 2: Distribution of 1253 students into 5 different clusters based on digital textbook engagement.

Cluster 2: Students in this cluster are indicative of Medium Engagement – Medium Performing learners. The 339 students in this cluster account for 27.06% of the total students and more than any other cluster. Although these students do not show the highest engagement they do not fall below the global average session time of 16.38. However, their average score of 83.03 is a little bit short of the global average score of 83.59.

Cluster 3: 13.65% of the total population of students fall into this group. They have the second lowest score across all 5 clusters. These students also have the least engagement metrics across the 5 clusters. Thus, their learning engagement is typical of Low Engagement – Low Performing students.

Cluster 4: The 284 students in this group have the second least session time and mean score across the 5 clusters. All the mean metrics of this group are below the global average. Just like cluster 3, this cluster is also indicative of Low Engagement – Low Performing students.

Cluster 5: These students have the highest session time across the 5 clusters but have the second highest score. Our initial expectation is that the students in the cluster with the highest mean session time will have the highest mean score. However, this did not hold true, and we suggest that this could be because of other metrics such as important markers. However, these students still perform above the global average and fall into the High Engagement – High Performing learnings.

4.2 Results of Prediction

The second analysis carried out on the data was the prediction of individual students' scores using the same features. The Fourteen features used in the clustering analysis were used here as attributes, except for the quiz scores which was used as the label to the predictive model. An XGBoost Regressor was used, with a test size of 0.2 of the instances. The Bar graph in Figure 3. Shows the performance of the prediction, using a comparison of the actual versus the predicted values of our model for 25 instances. The Root Mean Square Error of the prediction was 7.18 which is below 10% of the mean quiz scores (83.70), indicating that the model performed well for most instances.

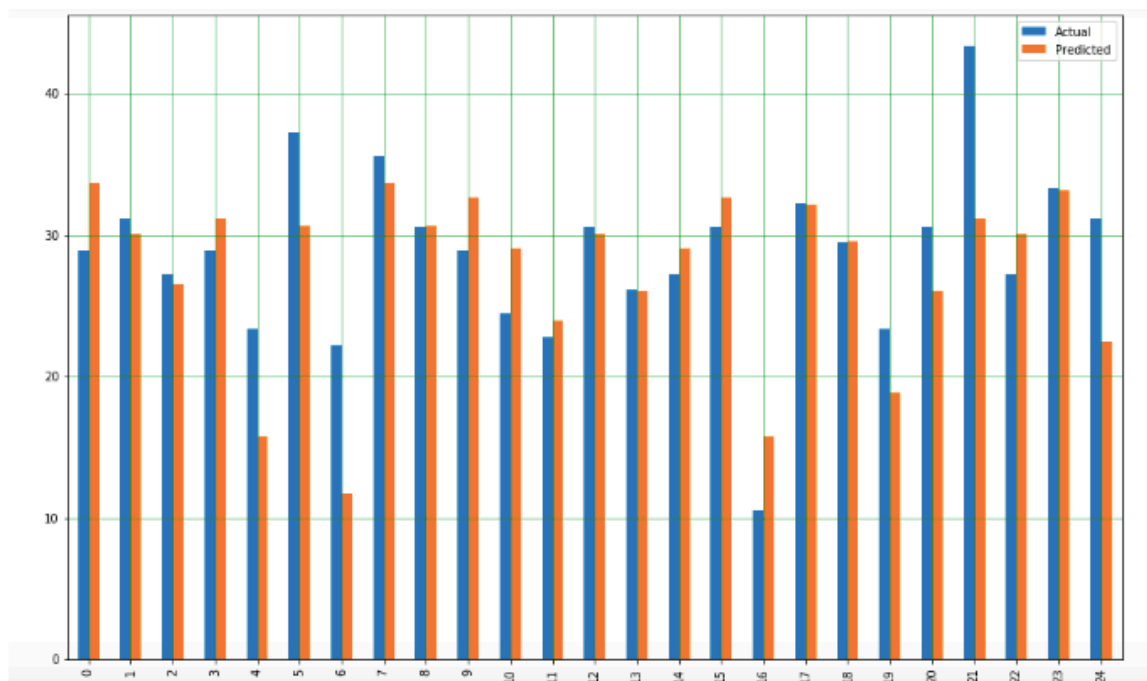


Figure 3: Graph showing Actual versus Predicted Quiz Scores for 25 Students.

5 CONCLUSION AND FUTURE WORK

In this paper, we analyzed students reading engagement in 10 different courses based on their activities on digital textbook reader, BookRoll. We identify 3 different analysis to be conducted towards predicting and improving students' performance. In the clustering analysis, 5 different clusters were obtained with each cluster indicative of either High Engagement – High Performing, Medium Engagement – Medium Performing or Low Engagement – Low Performing students. In the

concluding part of the work, predictions were made using the same features as used in the clustering analysis and the performance showed that the model used performed well. Future work will involve predicting the impact of pre-class and post-class activities on performance. As well as investigating a study group formation strategy for improving student performance.

ACKNOWLEDGEMENT

This research was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (S) Grant Number 16H06304.

REFERENCES

- Association of American Publishers (2014, June). US Publishing Industry annual survey reports \$27 Billion in net revenue, 2.6 billion units For 2013. Press Release Retrieved November 20, 2019, from: <http://www.publishers.org/press/138/>.
- Astin, A.W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel*. 25(4), 297–308.
- Atsushi, S., Fumiya, O., Hirokai, O. (2016). Browsing-Pattern Mining from e-book Logs with Non-negative Matrix Factorization. *9th International Conference on Educational Data Mining*. pp. 636-637, 2016, June.
- Atsushi, S., Takuro, O., Tsubasa M., Rin-ichiro, T. (2019). Learning Activity Analytics across Courses. *9th International Conference on Learning Analytics and Knowledge (LAK19)*. pp. 183-184.
- Atsushi, S., Kousuke, M., Yuta, T., Hirokai, O., Rin-Ichiro, T., Shin'ichi, K. (2019). Optimizing Assignment of Students to Courses based on Learning Activity Analytics. *12th International Conference on Educational Data Mining (EDM 2019)*. pp. 178-187.
- Bholowalia, P., & Kumar, A. (2014). EBK-means: A clustering technique based on elbow method and k-means in WSN. *International Journal of Computer Applications*, 105(9).
- Bossaller, J., & Kammer, J. (2014). Faculty views on eTextbooks: A narrative study. *College Teaching*, 62(2), 68–75.
- Flanagan, B., Ogata, H. (2017). Integration of Learning Analytics Research and Production Systems While Protecting Privacy. *In International Conference on Computers in Education*. pp.333-338.
- Longhurst, J. (2003). World history on the World Wide Web: A student satisfaction survey and a blinding flash of the obvious. *The history teacher*, 36(3), 343–357.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 14, pp. 281-297).
- Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-Book-based learning analytics in university education. *In International Conference on Computer in Education*. pp. 401-406.
- Siemens, G. (2010). 1st international conference on Learning Analytics and Knowledge 2011. Retrieved December 2, 2019, from: <https://tekri.athabascau.ca/analytics/>.

Learner's Performance Prediction based on Histogram of Actions during lecture

Takayoshi Yamashita

Tsubasa Hirakawa

Hironobu Fujiyoshi

Chubu University

{takayoshi, hirakawa, fujiyoshi}@isc.chubu.ac.jp

ABSTRACT: To predict the score as a learner's performance, we proposed the method based on recurrent neural network and histogram of actions during lecture. The amount of recorded actions of a learner is not fixed in a lecture. So, we create actions to histogram in each lecture. This histogram is considered the density of actions in temporal. We input the histogram to the recurrent neural network to predict the score. In the experiment, we compare with a neural network to show the importance of temporal information. As a result, the recurrent neural network achieves better prediction in both verification and testing data. Our method of RMSE is 5.93 points in the test dataset.

Keywords: Histogram of Actions, Recurrent Neural Network, score predictor

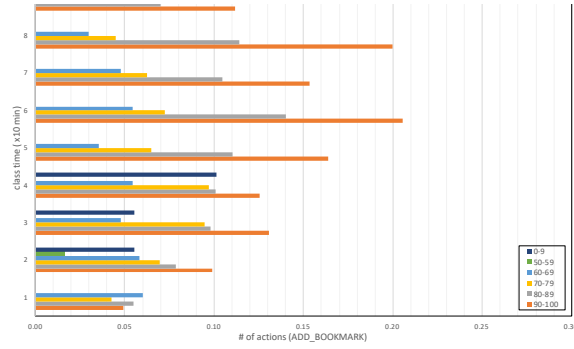
1 INTRODUCTION

In recent years, learning analytics (LA) that analyze collected learning activity data and improve education have attracted attention. It is possible to understand the learning situation by analyzing the action logs of electronic teaching materials in real-time. From these logs, the teacher can change the speed of lecture flexibly during the lecture. Besides, by predicting the results based on the action log, it is possible to grasp whether or not the student can acquire credit and to perform individual follow-up. In (Okubo 2017), it inputs the action log of each lecture to the Recurrent Neural Network (RNN) as a feature vector and predicts the final grade. In (Shimada 2016), it performs pattern mining by focusing on the relationship between students' browsing patterns of electronic teaching materials and their grades. In the case of (Okubo 2017), since the whole of the action log for each lecture is used as the feature vector, it could not capture the student's behavior during the lecture. The number of action logs of the student does not fix in one lecture. Therefore, we divide the action log of one lecture into a particular time and create a histogram of actions. It is possible to obtain the student's behavior during the lecture and the appropriate amount of the action log simultaneously. It feeds the histogram of actions to the RNN to consider the temporal behavior of the students during the lecture.

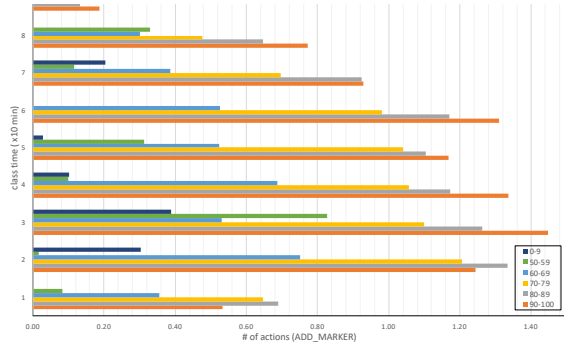
2 LEARNER'S ACTION DATA

The M2B is a learning support system to acquire the action log of electronic teaching materials during a lecture (Ogata 2017). It has several functions, and student action logs can be collected using the electronic teaching material system "BOOKROLL" (Flanagan 2017). Whenever a student operates

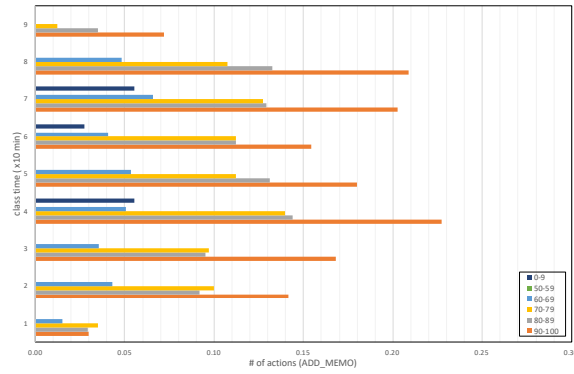
electronic teaching materials on BOOKROLL, the user ID, lecture material ID, page number, action type, and the time are automatically recorded as action logs. The action types include "OPEN", "CLOSE", "NEXT" indicating a click action to go to the next page, "PREVIOUS" indicating a click action to return to the previous page, and "ADD_MEMO" to make a note on the page, etc. There are 16 types of action logs.



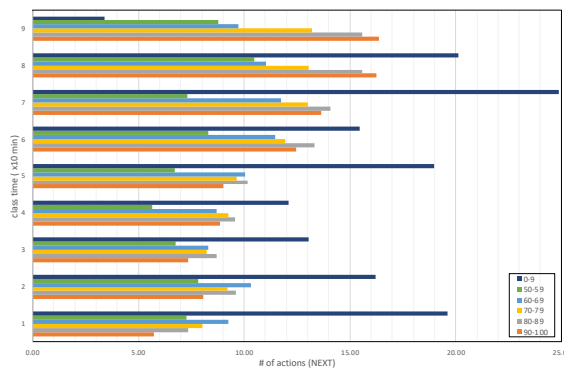
(a) action logs of ADD_BOOKMARK



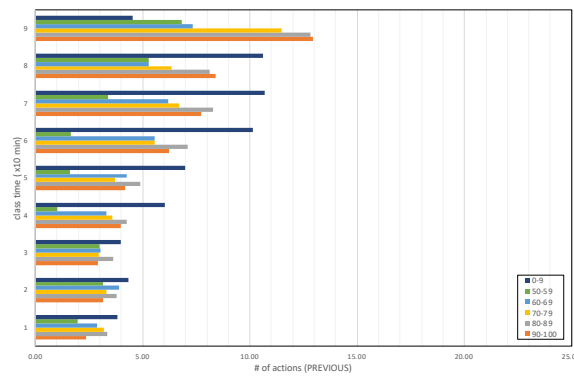
(b) action logs of ADD_MARKER



(c) action logs of ADD_MEMO



(d) action logs of NEXT



(e) action logs of PREVIOUS

Figure 1: action logs at each time unit (10min)

2.1 Provided dataset

In this study, we use the dataset collected by the system (Flanagan 2017). This dataset has two classes of the same lectures that are eight weeks, and the number of students is 1327. The learners may be absent sometimes. The total number of students in 8 weeks is 9602. For evaluation, there

are action logs for 396 students who attended the same lecture next year. Here, the action log is recorded as an action log when BOOKROLL is used not only during the lecture time but also outside the lecture time. In this study, it is limited only during lecture time. We analyze this action log in the next section.

2.2 Interaction analysis

The number of action logs during the lecture time varies. First, we divide the lecture time into 10-minute units, and we show the average number of each action log in Figure 1. The students with a higher score often use [ADD_BOOKMARK], as shown in Fig. 1 (a). [ADD_MARKER] in Fig. 1 (b) is also used by students with higher scores. There is a similar tendency for [ADD_MEMO] in Fig. 1 (c). [ADD_MARKER] is used more frequently than [ADD_BOOKMARK] and [ADD_MEMO]. Because it is a simple action of drawing a marker so that it can easily. For [NEXT] in Fig. 1 (d), the number of actions increases as the lecture time advances. However, the number of students with low scores often use at the beginning and the latter half of the lecture. Besides, the number of action decreases significantly in the end. It suggests that students with low scores are unable to see appropriate pages of electronic teaching materials. While it performs the quiz at the end of the lecture, it seems that the action logs of [NEXT] and [PREVIOUS] increase to see the page. The students with low scores could not answer the quiz because the number of actions decreases significantly. As a result, [NEXT] and [PREVIOUS] are also important actions as well as [ADD_MARKER] and [ADD_MEMO] in predicting the performance.

3 PROPOSED METHOD

The number of action logs of electronic teaching materials during the lecture time varies from student to student. Therefore, we perform the preprocessing that creates the histogram of actions as a feature vector. It inputs to the RNN as temporal data, and predict the score. Figure 2 shows an overview of the system.

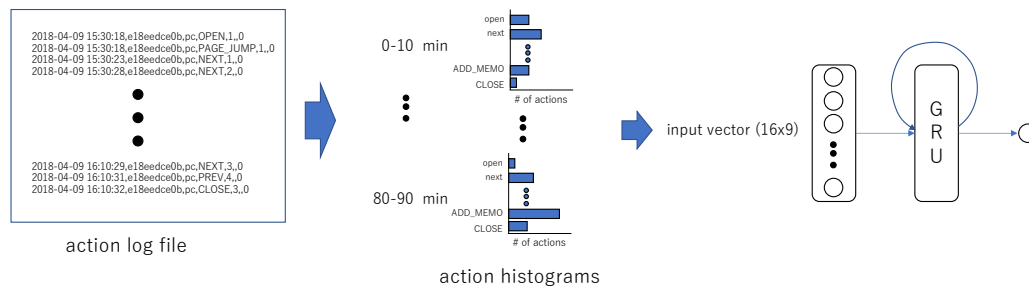


Figure 2: Overview of the proposed method

3.1 Preprocessing

First, we load the action log for each student. Then, we create a histogram of actions. It divides the action log for each lecture by 10 minutes. The time of one lecture is 90 minutes, and there are 16 types of action logs, so we obtain nine histograms with 16 bins. And since the number of lectures is 8, the total histograms are 56. Here, as the temporal data, the change of the histogram of actions in one lecture is essential, but the difference by the number of lectures does not need to be emphasized. Therefore, we input 9 histograms of actions in one lecture to the performance predictor in the next session.

3.2 Performance predictor

We employ the RNN as the performance predictor to receive temporal data. The simple Elman-type RNN cannot handle long-term temporal data. Therefore, we use the Gated Recurrent Unit (GRU) as that can process long-term temporal data. The architecture of the proposed network consists of 3 layers. We use the Mean Squared Error (MSE) as the error function.

4 EXPERIMENTS

We divide the provided data for training and verification. The provided data consists of two lectures held at different times. To evaluate an unknown lecture, we select a lecture with an earlier start time is used for training and a lecture with a later start time for verification. The number of training data is 933, and the number of verification data is 396. First, we compare the proposed method with a neural network (NN) to show the effectiveness of temporal information. Next, we compare the configuration of the RNN. Finally, we show the results for the test data.

4.1 Comparison with /without temporal process

To confirm the importance of temporal information, we compare our method with a neural network that consists of three fully connected layers. The number of units in the recurrent neural network and the hidden layer of the neural network are 192. The number of epochs during training is 100, the mini-batch size is 10, and we use Adam for the optimization method. At each epoch, we calculate the MSE for the verification data and choose the best model with the smallest MSE. Table 1 shows the comparison results. From Table 1, while the MSE of the neural network is 9.50, the MSE of the proposed method is 8.24. Although the neural network obtains a low MSE for the training data, it is almost 1.2 points higher for the verification data.

On the other hand, our method obtains the MSE for the training data, and the verification data are almost the same. Our method based on a recurrent neural network avoids over-fitting for training data. This result shows that temporal information is important for performance prediction.

Table 1: Comparison with/without temporal process.

network	RMSE (train)	MSE (val)
NN	8.19	9.50
RNN	8.72	8.24

4.2 Comparison of network architecture

We compare the network architectures of our method in this section. The parameters of training are the same as the previous section. We show the comparison results in Table 2. From this table, the RNN, which has 192 units in a hidden layer, achieves the best MSE for both training verification data.

Table 2: Comparison of the architecture of the proposed method.

network	MSE (train)	MSE (val)
RNN (96 units)	9.72	8.70
RNN (192 units)	8.72	8.24

4.3 Evaluation results in the test dataset

To evaluate the test dataset, we train the best network architecture with training and verification data. We show the result in the test dataset in Table 3. Our method achieves 5.94, using both training and verification data for training.

Table 3: Comparison of the test dataset.

network	RMSE
RNN (92 units, training only)	6.35
RNN (92 units, training + verification)	6.07
RNN (192 units, training + verification)	5.94

5 CONCLUSION

In this paper, we proposed a learner's performance prediction method based on the histogram of actions and recurrent neural network. To create a histogram of actions, we divide the lecture time and store the action at each corresponding bins of a histogram. We compare the recurrent neural network and neural network to show the importance of temporal information. As a result, the recurrent neural network achieves better performance than the neural network. We will consider more detail behavior of learners during lecture to improve the prediction performance.

REFERENCES

- Flanagan, B., Ogata, H. (2017). Integration of Learning Analytics Research and Production Systems While Protecting Privacy. In Proceedings of the 25th International Conference on Computers in Education (ICCE2017), pp.333-338.
- Ogata, H., Oi, M., Mohri, K., Okubo, F., Shimada, A., Yamada, M., Wang, J., Hirokawa, S. (2017). Learning Analytics for E-Book-Based Educational Big Data in Higher Education. In Smart Sensors at the IoT Frontier, pp.327-350, Springer, Cham.
- Shimada, A., Okubo, F., & Ogata, H. (2016). Browsing-Pattern Mining from e-Book Logs with Non-negative Matrix Factorization. In Proceedings of the 9th International Conference on Educational Data Mining (EDM2016), pp. 636-637.
- Okubo, F., Yamashita, T., Shimada, A., Konomi, S. (2017). Students' performance prediction using data of multiple courses by recurrent neural network. In Proceedings of the 25th International Conference on Computers in Education (ICCE2017), pp.439-444

Score Prediction Based on Page Feature Clustering

Ryusuke Murata, Tsubasa Minematsu, Atsushi Shimada

Kyushu University, Japan

murata@limu.ait.kyushu-u.ac.jp

ABSTRACT: In LAK 20 Data Challenge, we predict students performance based on their interaction with digital learning materials. In previous studies which analyzed on the relationship between reading behavior and students performance, researchers aggregate students activities lecture by lecture. In order to focus on more detail information, we aggregate students activities in each page in learning materials. Then, we cluster pages based on the reading time and the number of operations on the page. The percentage of the number of pages classified to each cluster represents features of students behaviors. We apply “LightGBM” to the features and predict students performance. LightGBM is one of decision tree algorithm and easy to understand what features contribute to prediction result. In our experiments, the result showed that the clusters mainly related to reading time were contributed to prediction.

Keywords: performance prediction, clickstream data, k-means algorithm, LightGBM

1 INTRODUCTION

With the recent increase in use of digital learning materials, the relationship between reading behavior and student performance has been actively studied. In LAK20 Data Challenge, we predicted students’ final quiz scores (0-100) through their course based on their interactions with digital learning material system “BookRoll”. In the previous study about the relationship between students performance and BookRoll usage, researchers aggregated students activities by the lecture (e.g. Askinadze, Liebeck & Conard, 2018). We focused on a little more detail information: students activities on each page in lecture materials. The page-based information can provide insights that are not obtained from the information in the entire lecture, such as how long students read each page and where they pay attention in lecture materials. In this paper, we clustered each page based on the reading time and the number of operations on the page, and student performance was predicted from their features represented by the cluster.

2 DATASET

In this data challenge, two datasets (Train and Predict) were provided. Train dataset and Predict dataset have 10 courses and 2 courses, respectively. Each course contains more than 100 students. All courses were consisted of eight lectures, and same subjects were taught to students. Courses in Train dataset was held one year before courses in Predict dataset. Train dataset has four types files, and Predict dataset has three types files: (1) click stream data collected by BookRoll, (2) the number of pages in each lecture material, (3) the schedule of each lecture, (4) the final total quiz score for each student (Train dataset only). The clickstream data records who manipulated what operation on what page. The recorded operations are opening / closing the lecture material, jumping to a

particular page, moving to next / previous page, adding or deleting bookmark / marker(highlight) / memo, clicking a link in current page, and searching words.

3 METHOD

Firstly, we aggregated the reading time and the number of each operation (bookmark, marker, memo, link click, and search words) on each page for each student in lectures. We used difference between the number of “add” and “delete” for bookmark, marker, and memo. We utilized the OpenLA (Murata, Minematsu & Shimada, 2020) library for these pre-processing. Then, we used k-means clustering for the pages in each lecture of all courses based on the reading time and the number of each operation. The number of clusters for each lecture was determined by the elbow method. The result is in Table 1, and it shows the number of clusters for each lecture is range of 8-10. Figure 1 is one of the clustering results of pages in a lecture. According to Figure 1, pages in cluster 0 were read for relatively short time, and there were not so many operations on the pages. The pages in cluster 1-5, and 7 were classified by the operation mainly manipulated on the pages. The pages in cluster 6 had characteristics of long stay and less operations. We guess that students left these pages open.

After the page clustering, we made a feature vector for each learner. First, we investigated the percentage of the number of pages in a lecture material assigned to each cluster. Therefore, N-dimensional feature vector was generated for a lecture material (N is the number of clusters for the lecture week). We generated feature vectors for all lecture week, and then, concatenated the feature vectors. Hence, 70-dimensional (total number of clusters in 8 lectures) vector was generated for representing each student’s learning feature. The feature vectors were used for training of machine learning model to predict students performance. The model used for prediction is “LightGBM” which is one of gradient boosting decision tree algorithm. Decision tree algorithm makes it easy to understand what features contribute to analysis result, and LightGBM is highly accurate in the decision tree algorithm. For more details about LightGBM, refer to the original paper (Ke et al., 2017).

Table 1: Lecture week and the number of clusters

Lecture week	1	2	3	4	5	6	7	8
The number of clusters	8	10	8	8	9	9	8	10

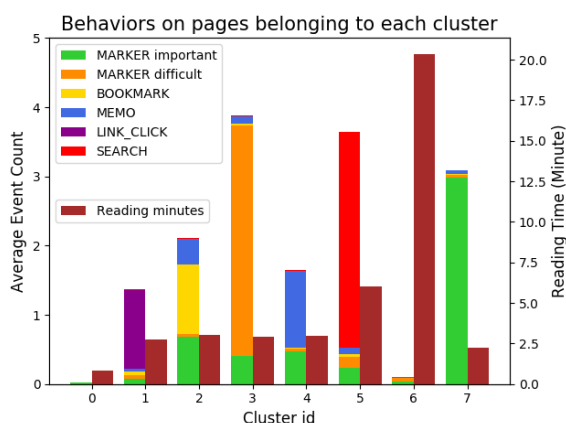


Figure 1: Clustering result in lecture week 1



Figure 2: Top 5 features contributed to predict

4 RESULT

We trained prediction model fifty times while separating students in Train dataset into train and validation data randomly. Then, we applied the models to feature vectors of students in Predict dataset, and the average of outputs is predicted score. The result of RMSE score was 5.61. Figure 2 shows the top 5 features contributed to predict. Feature importance in horizontal axis means how many the feature is used in fifty models on average. The first and fifth features (lectre1_cluster0 and lectuer1_cluster6) are equal to cluster id 0 and 6 in Figure 1. The second and third features (lecture7_cluster0 and lecture8_cluster0) are clusters including the pages which are read for short minutes and manipulated less operations, i.e. they are similar to cluster id 0 in Figure 1. The fifth feature lecture2_cluster7 is cluster including the pages which are left open for long minutes and manipulated less operations, i.e. it is similar to cluster id 6 in Figure1. This result shows that the reading time on each page mainly contributed to predict in our model.

5 CONCLUSION

In this data challenge, we proposed prediction method based on page clustering. Firstly, we aggregated students activities in each page in learning materials. Secondly, we classified pages based on the reading time and the number of operations. Thirdly, we generated the feature vectors from percentage of the number of pages assigned to each cluster. Finally, we inputted the feature vectors to prediction model. The result of RMSE score was 5.61. The contributed features to predict are clusters including the pages which are read short or long time and manipulated less operations. This result shows that the reading time on each page mainly contributed to predict in our model. Therefore, it may be possible to improve prediction by clustering reading time more detail. For example, although we clustered the reading time and the number of operations together in this data challenge, clustering them separately is worth trying.

ACKNOWLEDGEMENTS

This work was supported by JST AIP Grant Number JPMJCR19U1, and JSPS KAKENHI Grand Number JP18H04125, Japan.

REFERENCES

- Askinadze, A., Liebeck, M., & Conard, S. (2018). Predicting Student Test Performance based on Time Series Data of eBook Reader Behavior Using the Cluster-Distance Space Transformation. *In 26th International Conference on Computers in Education*.
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *In Advances in Neural Information Processing Systems* (pp. 3146-3154).
- Murata, R., Minematsu, T., Shimada, A. (2020). OpenLA : An Open-Source Library for e-Book Log Analytics, *The 2nd Workshop on Predicting Performance Based on the Analysis of Reading Behavior*.

Performance prediction using dimension-reduced activity data

Taisei Aoki

Kyushu University
aoki.taisei.516@s.kyushu-u.ac.jp

Yuya Kida

Kyushu University
kida.yuya.484@s.kyushu-u.ac.jp

Maiya Hori

Kyushu University
maiya-h@ieee.org

ABSTRACT: The proposed method uses Neural Network (NN)-based approach to predict performance of each student. In this challenge, however, it is difficult to build a complex model because the number of supervised data is not large. Therefore, the proposed method adopts an approach to construct simple NN model after pre-processing using Principal Component Analysis (PCA). When using student activity data as input, PCA is used to remove the correlation between input data. Here, the dimension reduction of the input data is realized by this processing. A NN-based model is constructed using the dimension-reduced data. As a result, it is possible to predict the score with high accuracy under the limitation of less supervised data.

Keywords: PCA, dimensionality reduction, Neural network

1 INTRODUCTION

Multiple regression analysis and time series analysis are performed as a technique for performing performance prediction. On the other hand, techniques based on deep learning have been able to demonstrate high performance. However, these methods require more data as the model becomes more complex. In this study, we solve this problem by reducing the dimensions of the input data.

2 METHOD

Figure 1 shows the flow of the proposed method. Data of the logged activity data from students' interactions can be acquired using the BookRoll system[1][2]. First, the obtained activity data is normalized and converted into multidimensional vectors for each student. The PCA is performed to remove the correlation between the data. By using the dimension-reduced data as an input of the neural network, the model is prevented from becoming complicated. Details of each are shown below.

2.1 DIMENSION REDUCTION WITH PCA

The activity data of over 1000 students in the course are acquired by BookRoll system. After vectorization of activity data for each student, the dimension of the data set is reduced by using principal component analysis (PCA), which is a method of multivariate analysis that synthesizes a variable called a principal component. Principal components represent the overall variation with a small number of uncorrelated variables. From the contribution ratio of the eigenvalue corresponding to each principal component, how much dimension reduction is performed is determined.

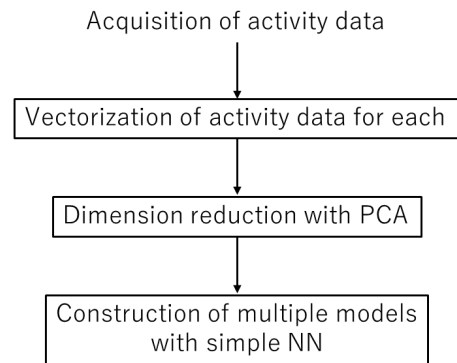


Figure 1 Flow of the proposed method

2.2 CONSTRUCTION OF MODEL USING THE DIMENSION-REDUCED DATA

In order to avoid complicating the model, we construct a NN model using the data with reduced dimensions. Performance prediction is performed by simple neural network regression with ReLU as an activation function.

3 EXPERIMENTS

The model was constructed as described above using the labeled dataset of 1000 students. The accuracy of the model was verified using a rest of the labeled dataset of 326 students as test data. Here, the cumulative contribution rate was calculated, and the original data with 120 dimensions was reduced to 65 dimensions. The cumulative contribution was 0.912. A simple neural network with one 10-dimensional hidden layer was used as a model. The RMSE value used as an evaluation measure was 4.81.

4 CONCLUSIONS

The proposed method uses NN-based approach to predict scores. The proposed method adopts an approach to construct a simple NN model after pre-processing using Principal Component Analysis (PCA). Our future work is to analyze in detail which factors are affecting the prediction.

REFERENCE

- [1] Brendan Flanagan, Hiroaki Ogata, Integration of Learning Analytics Research and Production Systems While Protecting Privacy, Proceedings of the 25th International Conference on Computers in Education (ICCE2017), pp.333-338, 2017.
- [2] Hiroaki Ogata, Misato Oi, Kousuke Mohri, Fumiya Okubo, Atsushi Shimada, Masanori Yamada, Jingyun Wang, and Sachio Hirokawa, Learning Analytics for E-Book-Based Educational Big Data in Higher Education, In Smart Sensors at the IoT Frontier, pp.327-350, Springer, Cham, 2017.

Predicting Student Exam Scores Based on Click-stream Level Data of Their Usage of an E-Book System

Makhlouf Jihed and Tsunenori Mine

Kyushu University

makhlouf_jihed@yahoo.fr, mine@ait.kyushu-u.ac.jp

ABSTRACT: In this paper we work with a dataset containing click-stream level data gathered from students' usage of a digital course-material-delivery system called "BookRoll". We build two different models to predict students' exam scores. We investigate their behavioral change when they use the system during the lesson and outside of the class time. Empirical data shows that a difference exists. We also propose a document-wise comparison of students' reading attitude to improve prediction performances. The experimental results show a marginal improvement of the model's predictions.

Keywords: E-Book System, Students' Performance, Reading Behaviors, Learning Analytics

1 INTRODUCTION

Educational institutions are constantly trying to improve students' learning experience, detect common students' behaviors and also predict their performances. Thanks to the continuously increasing adoption of educational software, big amounts of data can be gathered. The abundance of such diverse data, coming from different sources and in different forms and types, contributed to the growth of disciplines like the educational data mining and learning analytics. Actually, Learning analytics can improve learning practice by transforming the ways we support learning processes (Mavroudi et al., 2018), while educational data mining is more concerned with developing methods for exploring the unique types of data that come from educational settings (Baker, 2009).

As for today, different educational software systems exist in different forms and are being used in all levels of education. For example, ASSISTments (Heffernan et al., 2014), is an intelligent tutoring system for high school mathematics. Moreover, Learning Management Systems such as Moodle are being used in higher education. And, quite often they are a part of a more complex infrastructure serving for different systems. For example, the "BookRoll" digital teaching-material-delivery system allows teachers to upload lecture materials in a digital form which students can read anytime and anywhere (Flanagan & Ogata, 2018) (Flanagan & Ogata, 2017). The system provides students with different functionalities, such as markers, bookmarks and memos (Ogata, et al., 2015). In addition to the "BookRoll" system, the authors developed "SCROLL" which is a platform for Share and Reuse of Ubiquitous Learning Logs within an integrated system for learning analytics (Ogata, Li, Hou, & Yano, 2011).

The main objective of the "BookRoll" system is to provide the lecture content to students, but also it serves as a valuable source of data for learning analytics since it gathers students' usage data. Thanks to this system, the teachers are able to get feedback about the students' learning experience (Nakajima, Shinohara, & Tamura, 2013).

In an attempt to foster research into the analysis of students' interaction with digital textbooks, the team behind the "BookRoll" system and their collaborators organized a workshop in which they gave access to a dataset containing over 1000 students' usage logs¹. In the continuity of the first successful workshop, the organizers made another challenge where they opened a dataset from the "BookRoll" system.

Using the dataset, we investigate the differences between the students' reading behaviors during the lecture and outside of the lecture time. Moreover, we build two different models to predict students' exam scores. We introduce a document-wise comparison of students' behaviors. In fact, the content of a document has an influence on the behavior of students due to the material characteristics (length, topic, complexity of the topic...). Therefore, it is much fair to compare students' behaviors to their peers across the same documents. Experimental results show a slight improvement in the prediction performances for the document-based approach both in terms of the RMSE score and also the max error score.

The rest of the paper is organized as follows: Section 2 describes the dataset, the methodology adopted to generate the features and the genetic programming technique used to optimize the machine learning methods. In section 3, we provide the experimental results of the models' predictions. Finally, in the discussion section, we analyze the experimental results and elaborate some future improvements.

2 METHODOLOGY

2.1 Data Acquisition

In the present work, we use a dataset that was gathered from the E-Book reading system called "BookRoll". It is currently used in three different universities in Asia. More than 10,000 university students are using this E-Book system to access their course materials inside or outside of the classroom.

2.2 Initial Data Analysis

The dataset contains four types of files. Each type describes a different aspect of the gathered data. In the event stream file, each event is described by a set of information, such as the anonymized student ID, the document ID, the page number, the device (PC or Mobile), the timestamp, the action type and some other information that depends on the action type (Marker type, Memo length). The lecture is defined by an ID, start and end time, while the content is described by its number of pages and the lecture that uses it. When it comes to students, the dataset contains only their anonymized ID and their score in the respective course.

The main useful file is the action log data. Each action is characterized by its type (named 'operationname' in the dataset). It is a categorical feature having 15 possible values describing the

¹ <https://sites.google.com/view/lak19datachallenge>

types of actions that the student can perform. This is the feature that we used the most to generate even more features that help us gather more insights about students' reading behaviors.

Overall, the dataset contains almost 2 million rows of events. Each row describes an action made by a student while reading a document using the "BookRoll" system. The dataset consists of 10 courses taught across 80 lessons. Records contain the usage log data of 1326 students.

The first step was to cleanup and remove all duplicated data and also students for who we don't have their final scores. That left us with 1323 unique students in the dataset. We noticed that the same documents were used in similar lessons across all 10 courses; each course has 8 lessons. That makes it easier to investigate the document-wise comparison between students. However, to make sure we have a fair comparison, we need to verify how many documents the students had used.

In fact, as shown in Figure 1, each document has not been used by some students. However, using only Figure 1, we cannot make sure that students who did not use the documents are the same. Therefore, by using the Figure 2, we notice that there are only 6 students who used only one document, 16 students used 2 documents, 17 students used only 3 documents and we notice that the majority of students did use 6 documents or more.

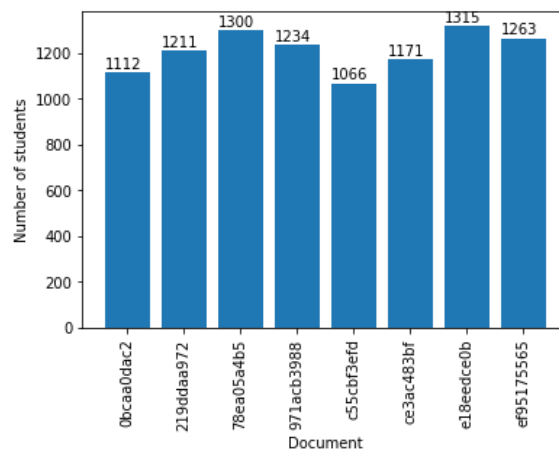


Figure 1: Number of students using each document

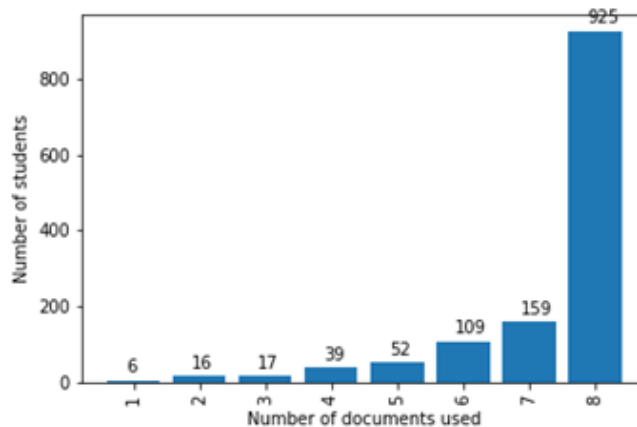


Figure 2: Number of students for each number of document usage

Another aspect of the dataset is the predicted variable. In fact, in this paper we predict students' exam scores. The range of the students' exam scores is from 50 to 100, while the mean score value is 83.681 and the standard deviation is 7.808. As shown in the histogram of Figure 3, the students' scores are highly concentrated between 80 and 95, hence the mean value is 83.

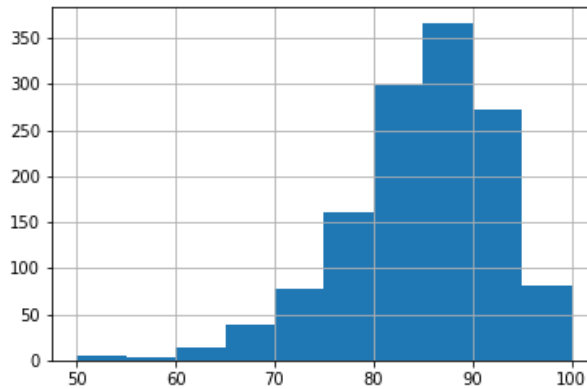


Figure 3: Histogram of the students' scores

2.3 Feature transformation

Since we are predicting students' exam scores, we have to change the granularity of our data from the event-stream level to the student level. While transforming the data, we investigate the students' behaviors. The first step is to separate between action types. For example, actions related to memos, such as adding, deleting or changing the memo, are aggregated together. We do the same thing to the bookmark, and the marker. However, the number of these actions is relatively low compared to the normal browsing actions which are pressing next page, or previous page, and that constitute the majority of actions. Such a situation creates a problem with the column variance being very low. Therefore, we decided to aggregate the marker, bookmark and memo actions altogether.

As explained in Table 1, the browsing actions are all actions related to displaying the contents of documents and the actions that allow the student to read through the document. The interaction actions are all actions that allow the student to act on the documents like the bookmark, the marker and the memo. Using the original categorical feature called "operationname", we can count these "Browsing" and "Interaction" actions separately, for each student. After that, we divide them by the total number of actions that were done by the respective student. Using this division, we calculate the ratio of "Browsing" and "Interaction" actions.

Table 1: Features composition

Feature name	Feature composition
Browsing	OPEN, CLOSE, NEXT, PREV, SEARCH, SEARCH_JUMP, PAGE_JUMP, LINK_CLICK
Interaction	{ADD, DELETE} BOOKMARK, BOOKMARK_JUMP, {ADD, DELETE} MARKER, {ADD, DELETE, CHANGE} MEMO

In addition to the “Browsing” and “Interaction” features we also calculate other features such as the action counts, the documents used, the total length of the memos and the ratio of difficult and important markers.

2.4 Students’ behaviors

At this point, we don’t apply and measure the features explained above. In fact, in the process of generating the new features, we make a distinction between the events happening during the lecture time, and the actions that were made outside of the lecture time. Therefore, we did measure all these features while separating between the actions that happened during the lecture and the ones outside of the lecture time.

As shown in Table 2, we have 7 features that we calculated during the class and also outside of the class. This generated an overall of 14 features. The {in, out}_lecture_browsing and {in,out}_lecture_interaction features are measured according to the composition in Table 1. We can easily notice that there is a statistically significant difference in the students’ usage when they are in the lesson and when they are not. Firstly, they don’t use the system a lot when they are not taking a class. It is clear from the number of actions and the number of documents used. Both of them are reduced drastically. Also, when they use the memo function, they deal with a lot shorter memos. However, the type of usage does not change a lot. In fact, the ratio of browsing actions is dropped slightly similarly to the interactive actions. Similar results appear to be happening for the marker usage. Overall, students do not use the system a lot when they are not in the lesson time.

Table 2: Difference between the student’s usage during the class and the usage outside of the class time

Feature name	Mean	Standard Deviation	T-test
in_lecture_actions_count	1219.33	685.46	48.211
out_lecture_actions_count	233.96	306.72	(p-value < 0.01)
in_lecture_docs_used	7.22	1.40	44.784
out_lecture_docs_used	3.86	2.33	(p-value < 0.01)
in_lecture_total_memo_length	822.91	2977.43	9.752
out_lecture_total_memo_length	23.34	167.81	(p-value < 0.01)
in_lecture_browsing	94.07	7.88	4.322
out_lecture_browsing	91.09	23.81	(p-value < 0.01)
in_lecture_interaction	5.84	7.45	9.587
out_lecture_interaction	3.08	7.37	(p-value < 0.01)
in_lecture_important	55.1	42.69	24.676
out_lecture_important	17.08	36.30	(p-value < 0.01)
in_lecture_difficult	23.42	33.79	12.939
out_lecture_difficult	8.31	25.74	(p-value < 0.01)

2.5 Document-wise comparison

Students' reading behaviors can be depending on the documents which they read. In fact, documents have different number of pages, and also different contents. For a defined course, the materials used in the introductory lessons are different from the materials used in more detailed and advanced topics of the same course. To avoid this situation, we propose to measure students' reading behavior for each document and compare it to their peer. Then we validate the effectiveness of this approach by comparing it to the baseline approach in terms of prediction performances.

As a measure of this approach, we use the z-score. It is a standardization method that indicates how many standard deviations that differs between a value and the mean. The formula to measure the z-score is as follow:


$$z = \frac{(X - \mu)}{\sigma}$$

where X is the value, μ is the mean, and σ is the standard deviation. Beside the standardization capability of the z-score, it is also frequently used as a ranking method. Therefore, using the z-score we rank the student's behavior compared to their peers in each document. Thus, the feature values consist now of the ranking of that student compared to other students that used the same document. Finally, to transform the dataset from the document level to the student level we take the mean value for each feature.

As shown in Figure 4, we use the z-score function to measure the ranking of the student features compared to the other students that used the same document. This is done by taking a sub-part of all students that used a particular document and we apply the z-score function to their features. In this way, we transform the features from normal values to "ranking" values.

2.6 Models building approaches

For the purpose of investigating the impact of document-wise comparison of students' behaviors we build two different models. In the baseline, we do not proceed to any specific change in the features. In the second approach we use the results of the document-based approach.



User_id	Contents_id	Feature1	Feature2
Student 1	Document 1	10.50	0.83
Student 1	Document 2	0	0.34
Student 2	Document 1	0	0.52
Student 2	Document 2	16.58	0.18
Student 3	Document 2	0	0.24
Student 4	Document 2	18.32	0.21
Student 4	Document 1	22.73	0.13

Figure 4: Comparing students' reading behaviors across the same documents using z-score

2.6.1 Splitting the data

In each approach, after the feature transformation, we proceed to splitting the data into train and test with a proportion of 3/4 training and 1/4 testing.

2.6.2 Feature Selection

For each approach, we separately apply a univariate feature selection. Some features were selected both times and some were different depending on the approach. Table 3 shows the results of the univariate feature selection process. The underlined features written in italic font are only selected in the baseline model. The features written in bold are only selected in the document-based model. The rest of the features are selected in both models.

Table 3: Feature Selection Results

<i>in_lecture_docs_used</i>	in_lecture_browsing	in_lecture_interaction
in_lecture_important	in_lecture_difficult	<i>out_lecture_total_memo_length</i>
out_lecture_important	out_lecture_difficult	in_lecture_total_memo_length
out_lecture_interaction		

2.7 Optimization and Genetic Programming

Since we have two different approaches, we want to compare them regardless of the machine learning technique that might be used to build the prediction models. Thus, for each approach, we want to find its prediction method with the best hyper parameters. To achieve this goal, we use genetic programming instead of the grid search method. Genetic programming can be described as a heuristic-based grid search. Therefore, it does not try all the possibilities which make it faster than the grid search method.

Briefly, genetic programming is a technique derived from genetic algorithms in which instructions are encoded into a population of genes. The goal is to evolve this population using genetic algorithm operators to constantly update the population until a predefined condition is met.

The most common ways of updating the population are to use two famous genetic operators called crossover and mutation. Crossover is used to diversify the research in the research space by taking some parts of the parent individuals and mixing them into the offspring. On the other hand, mutation is the process of updating only some part of an individual and it is used to maintain the actual diversity, in other words, intensify the research in a certain area of the research space. The population is evolving from one generation to another while keeping the fittest individuals in regard to one or many objectives. When using genetic programming for machine learning optimization, we use the model's prediction score as the objective function.

In our case, we use genetic programming by searching through a multitude of machine learning techniques and their respective hyper-parameters to find out which combination gives the best results. To achieve our goals, we use the python library TPOT (Olson, Bartley, Urbanowicz, & Moore, 2016). TPOT itself uses the scikit-learn python package and should be able to search through all machine learning methods that are implemented in it. Genetic Programming needs several hyper-parameters to be initialized in order to work properly.

Table 4 explores the principal hyper-parameters that we have to initialize. The Generations count is the number of iterations of the whole optimization process. A bigger number gives better results, but

also takes more time to finish. The Population size is the number of individuals which will evolve in each iteration. The offspring size is the number of individuals that are supposed to be generated from the previous population using the genetic algorithm operators. After executing the operators and generating the offspring, the individuals from the population and the offspring compete to survive and be part of the next population. When the individuals compete against each other, we only keep the fittest ones, meaning the individuals with the best score. The method used to measure the score is defined in the scoring hyper-parameters. We used the negative mean squared error as our scoring method. That means we only keep the individuals which have the closest score to zero. Mutation and Crossover rates are the probabilities of having respectively a Mutation or a Crossover operation to evolve one or more individuals. We set them to be 80% chance of having a mutation against 20% of having a crossover operation. Finally, to cross-validate our pipelines internally, we set the number of folds to 5.

Table 4: Genetic Programming Hyper Parameters

<i>Generations count</i>	<i>Population size</i>	<i>Offspring size</i>	<i>Scoring</i>
150	100	100	Negative MSE
<i>Mutation rate</i>	<i>Cross over rate</i>	<i>Internal Cross Validation</i>	
0.8	0.2	5-fold	

3 EXPERIMENTAL RESULTS

When the optimization phase finishes, it outputs the best machine learning pipeline and its best hyper-parameters that gave the best results. Since we executed the optimization phase for each approach separately, we gathered two different machine learning pipelines. Table 5 shows the results of the optimization phase and the best achieved scores. The Gradient Boosting (Friedman, 2001) method had the best RMSE of 7.70 for the baseline approach. For the document-based approach, the eXtreme Gradient Boosting algorithm (Chen, 2016) attained an RMSE of 7.61 that was the best during the optimization phase.

Table 5: Optimization Results

	<i>Baseline</i>	<i>Document-based</i>
ML Method	Gradient Boosting Regressor	eXtreme Gradient Boosting Regressor
Optimization RMSE	7.70	7.61

Once we obtained the machine learning techniques and their parameters, we use them to train models on the training set, then we test them on the held-out data. As shown in Table 6, prediction performances are improved in the document-based. In fact, the RMSE attain 7.19 in the document-based approach while the baseline has an RMSE of 7.31. Similarly, the max error in the document-based is lower compared to the baseline. However, the error variance in the baseline is higher, thus better, when compared to the document-based approach. In this case, both models' error variance is

not very good. In fact, the error variance in the baseline is about 0.017 while the error variance in the document-based approach is slightly worse, attaining 0.014.

Table 6: Validation results

	<i>Baseline</i>	<i>Document-based</i>
RMSE	7.31	7.19
Max error	34.05	31.29
Error Variance	0.017	0.014

4 DISCUSSION AND FUTURE WORK

In this paper we described how we used a dataset containing students' usage data of an E-Book reading system, called "BookRoll", to predict their performance. We investigated the difference in their behavior when there are in the lecture period and when they are not. Empirical results show that students when they are not in the lesson time, they don't use the system frequently. The whole number of actions performed drops drastically, but the ratio of browsing actions and interactive actions doesn't drop that much. Also, we noticed there is a statistically significant change in their behavior when they use the system outside of the lecture time compared to when they are in class.

In this work, we also predicted the students final scores using the log data. However, documents might have an influence on the behavior of the students. In fact, introductory lessons are different in contents and complexity from the more advanced topic. Therefore, it can affect the students' behavior. We avoided this case by comparing the students' behaviors to their peers when they use the exact same document. We call this approach, the document-based approach, and we validated it by building a prediction model and comparing it to a baseline model which takes no such document-based aggregation.

Experimental results suggest a minor improvement of the prediction performance of the document-based approach. In fact, the document-based model had better results in terms of RMSE and Max error, but failed to outperform the baseline model in regards to the error variance metric. However, the error variance is poor for both approaches. Having a low error variance might be caused the inability of the models to properly grasp the variance in the students' final scores.

One of the possible solutions is to add more features. In fact, in this study we did omit the time-related data that can actually encapsulate valuable information about the students' usage of the system. Taking into account the type of the action and the time-related data can generate valuable features. Since the document based approach showed a slight improvement of the prediction performances, we can use again with the time-related data as well as different courses

5 ACKNOWLEDGEMENT

This work is partially supported by JSPS KAKENHI No.JP16H02926, JP17H01843, JP18K18656 and JP19KK0257.

REFERENCES

- Baker, R. a. (2009, 01). The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*, 1, 3-17.
- Chen, T. a. (2016). XGBoost: A Scalable Tree Boosting System. *{Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining}* (pp. 785--794). San Francisco, California, USA: ACM.
- Flanagan, B., & Ogata, H. (2017). Integration of Learning Analytics Research and Production Systems While Protecting Privacy. *Proceedings of the 25th International Conference on Computers in Education (ICCE2017)*, (pp. 333-338).
- Flanagan, B., & Ogata, H. (2018). Learning Analytics Infrastructure for Seamless Learning. .
- Friedman, J. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, 29, 1189-1232.
- Heffernan, N. T. (2014, 12 1). The ASSISTments Ecosystem: Building a Platform that Brings Scientists and Teachers Together for Minimally Invasive Research on Human Learning and Teaching. *International Journal of Artificial Intelligence in Education*, 24, 470--497.
- Jihed, M. a. (2019). Investigating Reading Behaviors within Students Reading Sessions to Predict their Performance. International Workshop on Predicting Performance Based on the Analysis of Reading Behavior: A Data Challenge 2019.
- Mavroudi, O. V. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior*, 89, 98 - 110.
- Nakajima, T., Shinohara, S., & Tamura, Y. (2013). Typical Functions of e-Textbook, Implementation, and Compatibility Verification with Use of ePub3 Materials. *Procedia Computer Science*, (pp. 1344-1353).
- Ogata, H., Li, M., Hou, B. U.-B., & Yano, Y. (2011). SCROLL: supporting to share and reuse ubiquitous learning log in the context of language learning. *Research and Practice in Technology Enhanced Learning*.
- Ogata, H., Oi, M., Mohri, K., Okubo, F., Shimada, A., Yamada, M., . . . Hirokawa, S. (2017). Learning Analytics for E-Book-Based Educational Big Data in Higher Education. *Smart Sensors at the IoT Frontier* (pp. 327-350). Springer.
- Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-book-based learning analytics in University education. *Proceedings of the 23rd International Conference on Computers in Education (ICCE 2015)*, (pp. 401-406.).
- Olson, R. S., Bartley, N., Urbanowicz, R. J., & Moore, J. H. (2016). Evaluation of a Tree-based Pipeline Optimization Tool for Automating Data Science. *Proceedings of the Genetic and Evolutionary Computation Conference* (pp. 485-492). ACM.

A Picture-Book Recommender System for Extensive Reading on an E-Book System

Chifumi Nishioka^{1,*}, Sanae Fujita², Takashi Hattori², Tessei Kobayashi², Futoshi Naya²,
Hiroaki Ogata¹

¹ Kyoto University, ² NTT Communication Science Laboratories

* nishioka.chifumi.2c@kyoto-u.ac.jp

ABSTRACT: In this paper, we demonstrate a picture-book recommender system to promote extensive reading in English. Extensive reading refers to the independent reading of a large quantity of material for information or pleasure and is known to be effective for acquisition of a second language. The recommender system is implemented on an e-book system that shows digital learning materials (e.g., textbooks and slides) on student's device. Activities on the e-book system are recorded as learning logs. The recommender system suggests picture books in English based on contents of English textbooks. Specifically, we implement two recommendation strategies: (1) term-based recommendation and (2) grammar-based recommendation. In the future, we use learning logs and loan records of picture books to investigate the influence of picture books on acquisition of English and to enable personalized recommendations for each student.

Keywords: recommender system; language learning; e-book; learning analytics; extensive reading

1 INTRODUCTION

In these years, digital books (i.e., e-books) have been introduced to schools in different countries including Europe (Conrads et al., 2017) and Asia (Ogata et al., 2015). Activities on e-book systems are recorded as learning logs that are used for learning analytics. In this paper, we demonstrate a picture book recommender system on an e-book system for students to promote extensive reading in English. Although there are a lot of works of recommender systems for education and learning (Manouselis et al., 2011), to the best of my knowledge, there is no recommender system of resources for extensive reading. Extensive reading is defined as the independent reading of a large quantity of materials for information or pleasure (Renandya et al., 1998) and is reported to be effective for acquisition of a second language. According to Day and Bamford (2015), the primary purpose of extensive reading programs is to get students reading in the second language and liking it. Hafiz and Tudor (1989) reported that students prefer story books as reading materials. In addition, they indicated that shorter books place less strain on learners' concentration and are thus more likely to be picked up. Nishizawa et al. (2010) reported the effectiveness of a long-term extensive reading project, in which picture books were included in reading materials. Motivated by these works, we use picture books in English as reading materials for extensive reading. Although it is important provide a lot of picture books for resources of extensive reading, it is difficult to find a picture book that suits their levels and preferences. The recommender system aims to facilitate students to find picture books that match what they learn (i.e., words and grammar). As an e-book system, we use BookRoll (Ogata et al., 2015). BookRoll is a web application that shows digital

learning materials (e.g., textbooks and slides) on student's devices (e.g., tablet and laptop). Different activities on BookRoll are recorded as learning logs and will be used to investigate the influence of extensive reading on language learning.

2 PICTURE-BOOK RECOMMENDER SYSTEM

In this section, we describe the overview of the recommender system (Section 2.1) and recommendation strategies (Section 2.2).

2.1 Overview of the Recommender System

The recommender system is implemented on BookRoll (Nishioka & Ogata, 2018) and suggests picture books based on content shown on BookRoll. As shown in Figure 1 (left), we see a recommendation icon at the top-right corner, if there is at least one recommendation for the page. After clicking the recommendation icon, the recommendation panel is shown up at the right as illustrated in Figure 1 (right). It lists recommended picture books with metadata including author, title, and identification number. Each recommended picture book is attached a URL to a corresponding page in Google Books, where students can see the detailed information of the picture book. If a student is interested in a recommended picture book, she or he can find the picture book based on its identification number from the bookshelf of picture books installed in a school and borrow it. Clicks on the recommendation icon as well as recommended picture books are recorded as learning logs. In addition, loan records of picture books are also stored and will be used for learning analytics.

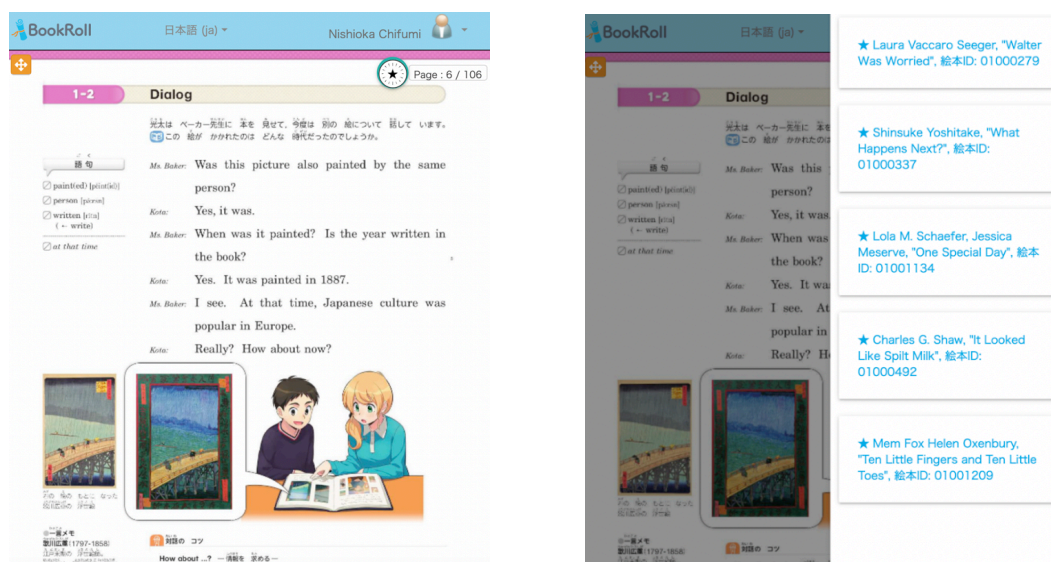


Figure 1: Interface of the recommender system. (Textbook: NEW HORIZON 3, Tokyo Shoseki, 2016)

2.2 Recommendation Strategies

The recommender system implements two recommendation strategies: (1) term-based recommendation and (2) grammar-based recommendation. We automatically extract texts from textbooks as well as picture books and use them for computing recommendations. As term-based recommendation strategy, we employ Term-Frequency Inversed Document Frequency (TF-IDF). As

grammar-based recommendation, recommendations are calculated along with the following procedures. We first manually assign grammar items to learn for each unit of English textbooks. Grammar items are listed by CEFR-J (2018). In contrast, we automatically detect which grammar items are used in each picture book. Then, we pick up picture books that uses the identical grammar items for each unit. Finally, among them, the recommended picture books are selected by TF-IDF. In the initial deployment, we employ the term-based recommendation strategy for even pages and the grammar-based recommendation strategy for odd pages. We compute term-based recommendations and grammar-based recommendations for each spread and unit, respectively.

3 CONCLUSION AND FUTURE DIRECTIONS

In this paper, we demonstrate a picture book recommender system to promote extensive reading for students. We deploy the recommender system in a junior high school in Japan since December, 2019. In the future, we would like to leverage learning logs recorded on BookRoll including clicks to recommendations as well as loan records of picture books to identify influence of recommendations and picture books on language learning. In addition, we would like to enable personalized recommendations of picture books using learning logs and loan records.

ACKNOWLEDGEMENTS

This work was supported by NEDO SIP 18102059-0, JSPS KAKENHI Grant Number 16H06304 and 18K13235, and JST AIP Grant Number JPMJCR19U1.

REFERENCES

- CEFR-J. (2018) Grammatical Item List Version: 20180315. http://cefr-j.org/sympo2018/pdf/CEFRJGP_GRAMMATICAL_ITEM_LIST.pdf#back
- Conrads, J., Rasmussen, M., Winters, N., Geniert, A., & Langer, L. (2017). Digital education policies in Europe and beyond. *JRC Science for Policy Report*.
- Day, D. D. & Bamford, J. (1998) *Extensive reading in the second language classroom*. Cambridge: Cambridge University Press.
- Hafiz, F. M., & Tudor, I. (1989). Extensive reading and the development of language skills. *ELT journal*, 43(1), 4-13. <https://doi.org/10.1093/elt/43.1.4>
- Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., & Koper, R. (2011). Recommender systems in technology enhanced learning. In *Recommender systems handbook* (pp. 387-415). Springer, Boston, MA. https://doi.org/10.1007/978-0-387-85820-3_12
- Nishioka, C., & Ogata, H. (2018). Research Paper Recommender System for University Students on the E-Book System. In *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries* (pp. 369-370). ACM. <https://doi.org/10.1145/3197026.3203882>
- Nishizawa, H., Yoshioka, T., & Fukada, M. (2010). The impact of a 4-year extensive reading program. In *JALT 2009 conference Proceedings* (pp. 632-640).
- Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-Book-based learning analytics in university education. In *International Conference on Computer in Education (ICCE 2015)* (pp. 401-406).
- Renandya, W. A., Jacobs, G. M., & Yu, V. W.-S. (1998). Extensive Reading in the Second Language Classroom. *RELJ Journal*, 29(2), 187–191. <https://doi.org/10.1177/003368829802900211>

What Activity Contributes to Academic Performance?

Tetsuya Shiino, Tsubasa Minematsu, Atsushi Shimada, Rin-ichiro Taniguchi

Kyushu University, Japan

shiino@limu.ait.kyushu-u.ac.jp; minematsu@limu.ait.kyushu-u.ac.jp;

atsushi@ait.kyushu-u.ac.jp; rin@kyudai.jp

ABSTRACT: we propose an approach designed to analyze relationships between learning activities from clickstream data and academic grades, through a neural network-based model. The dataset used in this study was collected from an e-book system, during a total of eight 90 min face-to-face lectures. We analyzed a total of 372,197 e-book learning logs from 162 students. We then modeled the relationships between students' grades and eleven types of learning activities, based on a neural network. By using Layer-wise Relevance Propagation (LRP), we investigated the importance of students' learning activity in relation to academic grades. In this paper, we explain our analytics strategy and report the primal results.

Keywords: Learning Activity Analytics, Neural Network, Layer-wise Relevance Propagation

1 INTRODUCTION

Recently, a large number of learning activity logs have been collected from digital learning environments, such as massive open online courses (MOOCs) and M2B systems in Kyushu University (Ogata, et al., 2015). These learning logs have been analyzed (Gitinabard 2018) (Park 2017). One of the main purposes of analyzing these logs is to understand how grades and learning activities are related because it can be a cue to support learning. For example, (Shimada 2016) conducted browsing patterns mining from learning logs and discussed the relationship with grades. Most of the traditional approaches are based on liner analysis, so that they sometimes fail to grasp reasonable relationship between learning activities and grades. In this study, we proposed an approach based on non-liner method using a neural network-based model. Our method shows the relevance of each learning activity for grades based on Layer-wise Relevance Propagation (LRP). In our experiment, we made a model that expresses the relationships between grades and learning logs in a course and visualized them as preliminary experimental results.

2 METHOD

Our method makes a single model for individual course, and analyze the relationship between learning behavior patterns and academic performance.

2.1 Learning log collection

Students' learning logs were collected from 162 students who attended a cyber-security course consisting of eight 90-minutes face-to-face lectures. The total of the operation logs collected from the courses was 372,197. In this course, the students used an e-book system which recorded the operation logs made by students. There are about 15 operation types recorded as logs. In this study, we focused on the specific operation types that are performed relatively frequently and Table 1 shows the

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

operation types we focused on. In addition, a quiz with a maximum score of 10 was performed at the end of each lecture. We used the average quiz score of each student as their grade.

Table 1: List of operation types

Operation Types	Actions
NEXT, PREV	Went to the next/previous page
ADD MARKER	Added a marker to the current page
ADD BOOKMARK, BOOKMARK JUMP	Added a bookmark/Jumped to a bookmark page
ADD MEMO	Added a memo in the current page
SEARCH	Searched for something within the e-book
GET IT, NOT GET IT	Clicked on the response button regarding whether the Student understands the current page
CLICK RECOMMENDATION	Access the recommended web page related to the current page

2.2 Analyzing relationships between learning logs and students' grades

We modeled the relationships between students' grades and their learning logs through a multilayer perceptron(MLP) which was a type of neural network. Our neural network predicted a student's grade from his/her learning logs. In this study, 162 students were divided into three groups: "good," "poor," and "others" according to their average quiz scores. Their numbers were 20, 119 and 23, respectively.

We designed 11 features related to students' learning activities as an input of the neural network. These features are simple to make it easier to understand the results of LRP. The nine features were computed by aggregating the number of each operation by a student. The rest represent the number of memo characters and pre-study time. We computed the 11 features week by week over 8 weeks, and then we obtained 88-dimensional feature vectors for each student. In this study, we intentionally made an overfitting model in order to investigate the contribution of each feature to the grades within a single course. The trained model achieved 100% accuracy for the training data.

In order to investigate the contribution of students' leaning activity to their academic grades, we used Layer-wise Relevance Propagation (LRP) (Bach 2015). LRP provides relevance scores for each element of an input feature vector. The relevance score represents the contribution of the element to the output inferred from the input vector. In the analysis of our neural network, we could consider that features with a large relevance score is the cause of the result.

3 RESULTS & DISCUSSION

We computed the relevance scores of the two groups "good" and "poor" and averaged the relevance scores respectively. By visualizing this result, we Investigate the cause of specific grades. Figure 1 shows the results of the average relevance scores for each group. For easy understanding, we represented relevance scores corresponding to 88-dimensional feature vectors as an 11x8 heat map. The horizontal axis is the feature type, while the vertical axis corresponds to each week. Comparing the two heat maps, we observed that the features contributing to each output were different. When focusing on the heat map of the poor grade students, features with large relevance scores were observed in the first and second lectures. For example, "NEXT" and "PREV" and "GET IT" in the first week and "NEXT" and "PREV" in the second week. This indicates that students with poor grades may not have been performing well from the beginning of the lecture. On the other hand, focusing on the

heat map of the good grade students, the relevance scores of “NEXT” and “PREV” in the fifth week is large. In fact, the lecture contents in the fifth week were more difficult than those of the other weeks, meaning students with good grades may have read the lecture material carefully, resulting in better grades. As a way to use this result, we consider advice to the students after the course. For example, let students who have poor grades know the difference from those who have good grades to encourage future learning improvement. However, it is necessary to carefully consider what kind of advice to give from this result. In the future, we plan to make a feedback system based on the LRP for students.

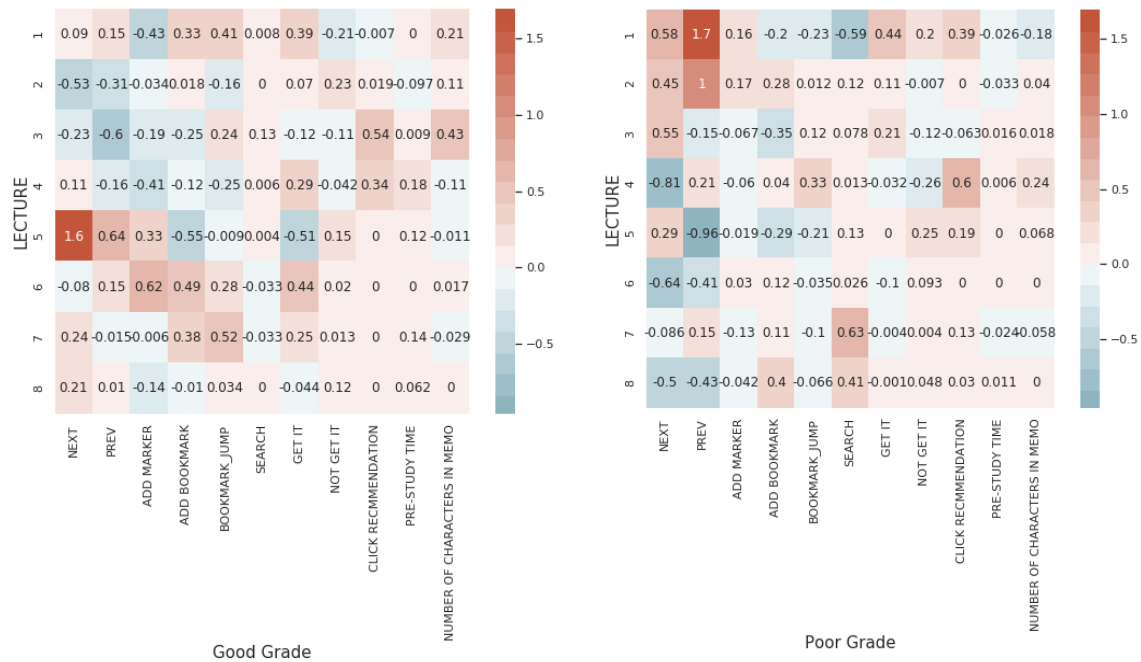


Figure 1: The Results of Applying LRP to the Analytics Model

ACKNOWLEDGEMENT

This work was supported by JSPS KAKENHI Grant Number 19H04226, JP18H04125, and JST AIP Grant Number JPMJCR19U1, Japan.

REFERENCES

- Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., Yamada, M. (2015). E-Book-based learning analytics in university education. *Proceedings of the 23rd International Conference on Computer in Education (ICCE 2015)* pp.401-406.
- Gitinabard, N., Khoshnevisan, F., Lynch, C. F., & Wang, E. Y. (2018). Your Actions or Your Associates? Predicting Certification and Dropout in MOOCs with Behavioral and Social Features. *arXiv preprint arXiv:1809.00052*.
- J. Park, K. Denaro, F. Rodriguez, P. Smyth, M. Warschauer. (2017). Detecting Changes in Student Behavior from Clickstream Data. *LAK'17*.
- A. Shimada, F. Okubo, H. Ogata (2016). Browsing-Pattern Mining from e-Book Logs with Non-negative Matrix Factorization. *the 9th International Conference on Educational Data Mining*, 636–637.
- Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K. R., & Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS one*, 10(7), e0130140.

Automatic Retrieval of Learning Contents Related to Quizzes for Supporting Students' Enhanced Reviews

Takashi Ishikawa, Tsubasa Minematsu, Atsushi Shimada, Rin-ichiro Taniguchi
Kyushu University, Japan
ishikawa@limu.ait.kyushu-u.ac.jp

ABSTRACT: We propose a system that suggests e-book pages to students in relation to quizzes. The related pages are automatically explored by analyzing similarities between contents of e-books and quiz questions. In this paper, we introduce the proposed analytics strategy based on two techniques, “TF-IDF” and “Word2Vec.” We evaluated the proposed method by lecture materials and their related quizzes in cyber security courses, and found a high possibility of enhancing students’ reviewing abilities.

Keywords: Learning Analytics, Related Page Mining, Natural Language Processing

1 INTRODUCTION

In recent times, digital learning environments are often utilized, for example, to provide lecture materials via an e-book system (Ogata, 2015), or for conducting quizzes and/or collecting learners’ reports via a learning management system, and so on. A good collaboration between these systems is expected to realize an advanced supporting strategy for enhancing students’ learning. The target of this study is to support a review process of a student after lectures. At the end of most lectures, students are asked to answer short quizzes on its contents to check their understanding; unfortunately, some students fail due to a lack of understanding. Such students are requested to perform a review of the lecture materials. However, some students have yet another difficulty in performing review processes due to less ability or low motivation. For supporting such students, one of the possibility is to suggest some specific pages related to the quizzes closely. This solution requires teachers to make a review material, but it is not realistic because most teachers are too busy to make review materials. To solve this problem, we propose a supporting system that automatically identifies e-book contents related to the quiz in which the student failed, and suggests relevant pages for review (Figure 1). We introduce our proposed analytics methodology, followed by preliminary experimental results.

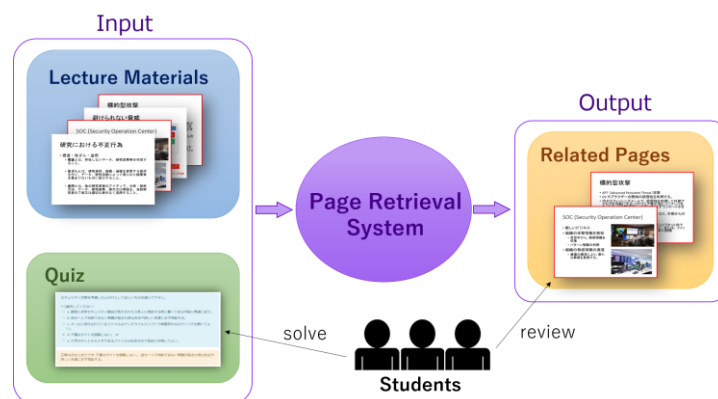


Figure 1: System overview

2 METHOD

Our proposed system uses a set of lecture materials and a quiz as a query in order to retrieve related pages from the lecture materials. Each quiz question consists of a statement and multiple-choice answers. A simple key page retrieval for answering a quiz is to find text pages from the materials containing words in the quiz; however, this page retrieval method cannot work effectively when key pages do not match words or phrases exactly. In order to solve this problem, we measure semantic similarity between words in text pages and quizzes. Word2Vec (Mikolov, 2013) is one of the useful methodologies to evaluate the similarities between words, so that we used Word2Vec method in our implementation.

2.1 Feature Extraction from Lecture Materials

The page feature is represented by 200-dimensional vector, which is generated by Word2Vec method. We utilized an existing Word2Vec model¹ made from Japanese Wikipedia (Suzuki, 2018). First, we extract words in each page of lecture material. Then we calculated TF-IDF (Ramos, 2003) value of each word in each page. In each page, each extracted word is input to the Word2Vec model, and its corresponding 200-dimensional vector is generated. Since a page has multiple words, the generated 200-dimensional vectors are weighted by TF-IDF value and averaged, and finally, all pages are represented by 200-dimensional vector.

2.2 Feature Extraction from Question Texts

We made a feature vector (200-dimensional vector) of a quiz text as well as the lecture materials. We calculated the mean vector of the words in the quiz text, and used the vector as a query of page retrieval. The similarity between the quiz text and page text was measured by cosine similarity. The system can identify page(s) containing not only the word in the question text but also words highly related to the target word.

3 EXPERIMENTS AND FUTURE WORK

A learning supporter, who understood the contents, selected a set of relevant pages. We used the pages as ground truth. Materials from seven lectures on cyber security courses and related quizzes were used for the experiment. The lecture materials used for each lecture had 27 to 81 pages, and each lecture had 5 to 10 multiple choice quizzes (total 46 quizzes). By inputting text data of lecture materials and quizzes, we calculated the similarities between each page vector and each quiz vector. Then, for each quiz, we sorted the retrieved pages in order of the similarity.

We evaluated the accuracy by Cumulative Matching Characteristics (CMC) curves using three query types such as “quiz statement only,” “quiz statement and multiple choices,” and “quiz statement and

¹ Word2Vec made by the laboratory of Prof. Inui and associate Prof. Suzuki in Tohoku University from <https://github.com/singletonue/WikiEntVec/releases>

correct answer.” As evinced in Figure 2, results were totally better when quiz statements and corresponding multiple choice or correct answers were used as a query. Especially in higher rank (rank 1 ~ 5), the system achieved higher accuracy rates when using correct answers as a part of the query than when using multiple choices. Findings indicate that better retrieval results are obtained when considering not only quiz statements but also correct answer statements as a query. We believe that the retrieved pages would support the review of lecture materials and help understanding of the contents.

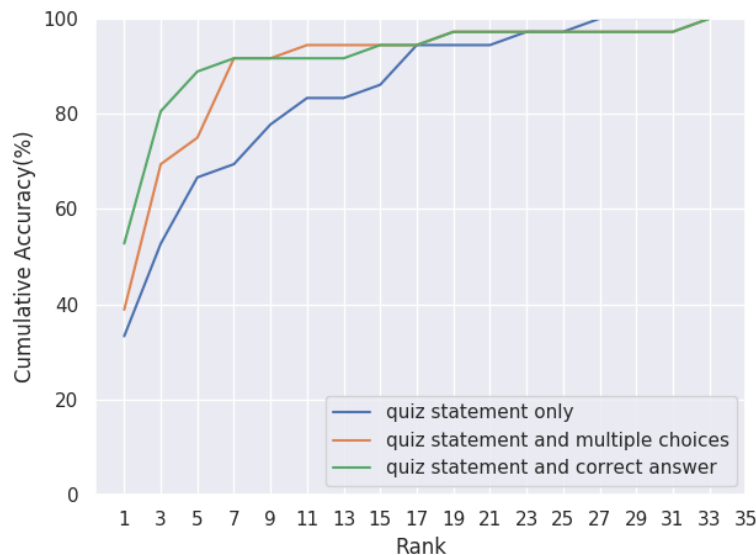


Figure 2: CMC Curve Graphs

In our future study, we will develop a system to provide retrieved pages based on the quiz result of each student. In addition, we will investigate the effectiveness of the system from the viewpoint of how much the review process can lead to a deeper understanding of lecture content.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grand Number 19H01716, JP18H04125, and JST AIP Grant Number JPMJCR19U1, Japan.

REFERENCES

- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119).
- Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., Yamada, M. (2015). E-Book-based learning analytics in university education. *Proceedings of the 23rd International Conference on Computer in Education (ICCE 2015)* pp.401-406.
- Suzuki, M., Matsuda, K., Sekine, S., Okazaki, N., & Inui, K. (2018). A joint neural model for fine-grained named entity classification of wikipedia articles. *IEICE Transactions on Information and Systems*, 101(1), 73-81.
- Ramos, J. (2003, December). Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning* (Vol. 242, pp. 133-142).

Generating individual advice corresponding to the learning level by analyzing learning behaviors

Taisei Aoki

Kyushu University

aoki.taisei.516@s.kyushu-u.ac.jp

Maiya Hori

Kyushu University

maiya-h@ieee.org

Atsushi Shimada

Kyushu University

atsushi@ait.kyushu-u.ac.jp

ABSTRACT: We propose a system that generates learning advice to efficiently improve grades of students by analyzing the learning behavior of each student. Existing systems usually recommend increasing the amount of learning time for students. However, grades do not improve as much as desired because the advice is not specific for each student. Our system attempts to overcome this problem by considering the learning level of each student. The advice for each student is generated by comparing the learning behavior with that of students who have a little higher learning level. In the experiments conducted, advice for each learning level was generated by analyzing the learning behavior of the students.

Keywords : e-learning, learning advice, learning level

1 INTRODUCTION

The e-learning system is composed of electronic materials and the learning support system (LSS) used for education. G. Hwang et al. [1] analyze Learning behaviors of each student acquired from the LSS. Q. Hu and H. Rangwala [2] predict that grades of the students from the learning behavior will help to assess the students that were likely to fail. Based on this prediction, such students could be accordingly advised to increase their study time. In specialized courses, however, it would not be possible to ensure that grades improve efficiently because the advice given is not specific for each student. Therefore, the focus must be on the learning level of each student, individually. Thus, suitable advice can be given for each student's particular learning level. As a result, the system can help students to keep on learning.

2 METHODS

Figure 1 shows the overview of the proposed system. This system can provide the advice that is suitable for the learning levels of each student. The system consists of following processes: (i) prediction of the final test score using features about learning behavior, (ii) estimation of the learning level, and (iii) generation of advice. By repeating these processes, individual advice is given at each lecture every time. As a result, the grades of the students will show improvement.

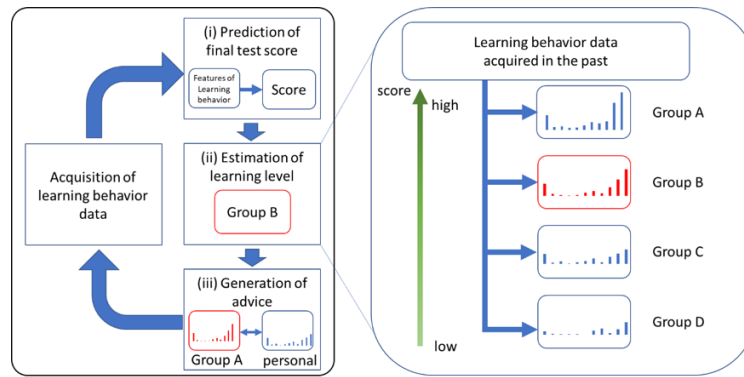


Figure 1: Overview of the proposed system

2.1 Prediction of the final test score during lecture term

The final test score for each student is predicted every time a lecture is given. Deep neural network (DNN) is used for the prediction of the final test score by using features of learning behavior of each student as an input. The DNN model is built in advance using the data of students who attended the previous year. Here, it is important to predict the final test score before the lecture term ends. This is important for improving performance by giving learning advice to students who are likely to have poor final scores.

2.2 Estimation of the learning level for each student

It is important to estimate each student's learning level to give appropriate advice to individual students. The learning level for each student is judged from the criteria decided from the predicted final test score. The criteria is decided using the final test score of students who took the same lectures in the previous year. The proposed method applies such an approach because the final test scores were not a simple normal distribution.

2.3 Generation of the advice corresponding to the learning level

The advice for each student is generated by comparing the learning behavior with that of students who have a little higher learning level. The important thing here is to refer the behavior of students whose learning level is not far apart. This makes each student can get advice that is suitable for their particular learning level. As a result, the grades of the students will show improvement.

3 EXPERIMENT

We conducted experiments using real data acquired from the lecture on the circuit theory as a specialized subject. Specialized subjects tend to have high knowledge relevance in previous and next lectures because the content learned at each lecture is not independent. The number of subjects were 76. The learning behavior such as the number of times the student operates the materials during a class, the number of times the student adds markers, the number of times the student attends class, and the number of times the student pushes "NOTGETIT" button when the student does not understand the page, is analyzed and 11 features of leaning behavior were generated. The final test score for each student is predicted by using features of learning behavior of each student. The students are classified into four groups: Group A, B, C and D according to their predicted final test score by a k-means algorithm as shown in Figure 1. Group A has the highest grades while group D

scores are the lowest. The advice for each student were generated by comparing the learning behavior with that of students who have a little higher learning level. Figure 2 shows an example of advice presented to 3 students belonging to Group B. The part of the behavior that is missing for each student is highlighted in red. As an example with ID648, the student can see that the number of slide operations is small, the number of slides that are not recognized is small, and the quiz score and attendance are low as compared with Group A students. From the results of other ID students the advice varies from person to person in the same group. Each student can understand what behavior leads to improvement of the grade by seeing this.

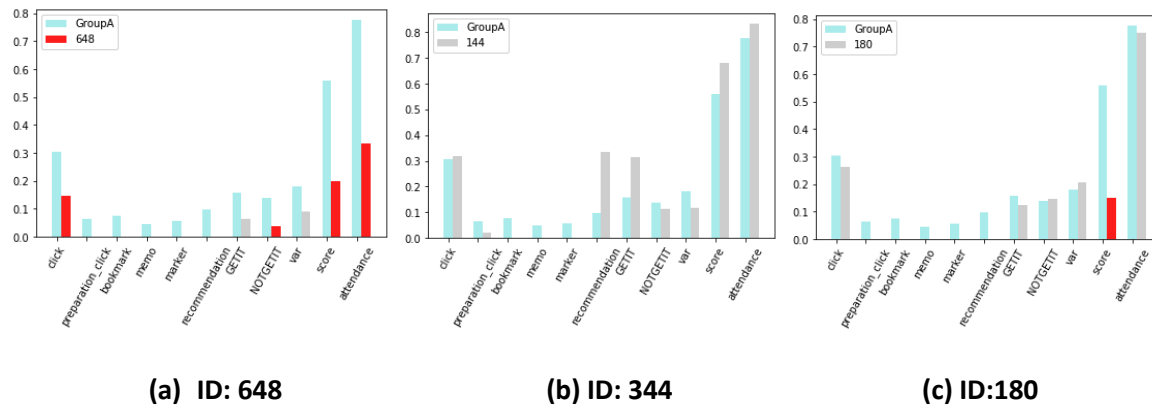


Figure 2: Individual advice for each student in the same learning level group. These are advices presented to 3 students belonging to Group B. The part of the behavior that is missing for each student is highlighted in red.

4 CONCLUSIONS

Thus, the proposed system is one that generates learning advice that will help to efficiently improve grades of students by analyzing the learning behavior of each student. In the experiments, it was found that the advice given to each student should be generated by considering their individual learning level. Our future work is to confirm the improvement of the grades of the students by using generated advice.

ACKNOWLEDGEMENTS

This work was supported by JST AIP Grant Number JPMJCR19U1, and JSPS KAKENHI Grand Number JP18H04125, Japan.

REFERENCES

- [1] Gwo-Jen Hwang, Hui-Chun Chu, Chengjiu Yin. Objectives, Methodologies and Research Issues of Learning Analytics. *Interactive Learning Environments*, Volume 25, Issue 2, 1 March 2017, 143-146.
- [2] Q. Hu and H. Rangwala. Reliable Deep Grade Prediction with Uncertainty Estimation. *In proc. of International Learning Analytics and Knowledge (LAK) Conference 2019*, 76-85.

Evaluating the Accuracy of Real-time Learning Analytics in Student Activities

Takuro Owatari, Atsushi Shimada, Tsubasa Minematsu, Rin-ichiro Taniguchi

Kyushu University, Japan

owatari@limu.ait.kyushu-u.ac.jp

ABSTRACT: A real-time learning analytics system allows teachers to immediately grasp student learning status and flexibly control their lectures. In this paper, we investigate the reliability of our real-time learning analytics system focusing on following two points: 1) Reliability of each student's page browsing transitions 2). Appropriate calculation cycles for real-time system. The experiments were performed using click-stream data from two face-to-face style lecture courses which consist of eight 90-minute lectures respectively. A total of 329 students attended the lectures, and approximately 703,000 click events were recorded via an e-book system. As a result, the system demonstrated enough reliability, and the calculation cycle of 30 seconds seems effective for real-time processing.

Keywords: real-time analytics, real-time system, learning behavior, e-Book event stream, evaluation

1 INTRODUCTION

Real-time learning analytics have tremendous potential to immediately grasp and improve the learning behavior of students. Particularly in classroom settings, real-time analytics can provide teachers with information about whether students are following the lectures, which textbook pages students are viewing or notable changes of students' activity (Shimada 2018b), allowing teachers the flexibility to change the pace at which the lecture is progressing in accordance with students' behaviors. An important function of a real-time feedback system is to grasp students' browsing behaviors during lectures (Shimada 2018a), however, research on the accuracy of the system's ability to reflect learners' situations in real time and analysis of the validity of per-minute processing for browsing behavior has not been conducted. Therefore, as one of the verifications required for using that real-time system (Shimada 2018a), we investigate the reliability focusing on two aspects: 1) Reliability of tracking each student's page browsing transitions by using collected click-stream data. In an actual environment, event logs cannot always be entirely recorded due to system failures, problems of implementation, usage environment, etc. 2) Appropriate calculation cycles for real-time systems. In general, calculation cost is proportional to the size of the dataset.

2 METHODOLOGY

Real-time systems summarize each student's event logs for a particular period of time to track browsing behaviors, such as duration of browsing. This summarization is performed at predefined intervals, such as once every minute, etc. In real-time systems, student's event logs get collected sequentially, and contain event timestamps and page numbers specifying where the events occurred.

In this study, we analyze differences between sequential processing results acquired by a real-time system and batch processing results acquired by collective data after the lecture.

In order to accurately track browsing transition, a real-time system should consistently collect event logs. Therefore, first, we identify inconsistent logs (which should ideally never occur as they should continue on from previous logs) as we follow each student's sequential logs in the time series. As an example, if the previous event was logged on page 1 and the operation was GO TO NEXT PAGE, the next event logged should occur on page 2. If, in this example, the next log occurs on a page other than page 2, it will be considered to be an inconsistent log. With real-time systems, there is no guarantee that the next log will occur on page 2, and, therefore, will sometimes result in an inconsistent log being observed after the interval is calculated. In addition, for each student, we computed unreliable browsing time when browsing tracking malfunctioned due to the occurrence of an inconsistent log in the real-time system.

Next, we verified the summarization by per-minute processing. Let t be the processing interval of a summarization. In addition, we defined missing browsing time as the total browsing time except browsing time of the page that the student browsed for the longest time in each processing cycle. Here, missing browsing time indicates information omitted by summarization. Missing browsing time of each student during the lecture is calculated for every processing cycle while varying the processing interval t . At that time, the processing time is also measured.

3 EXPERIMENTS AND RESULTS

The experiments were performed using click-stream data of two face-to-face style lecture courses consisting of eight 90-minute lectures each. In all, 329 students attended the lectures and approximately 703,000 click events were recorded via the e-book system (Ogata, 2015).

Figure 1 shows the result of unreliable browsing time. The blue bars indicate the average unreliable browsing time of each student in each lecture. The highest average is approximately 142 seconds, and this value is 2.6% of 90 minutes. Additionally, the red bars depict the average number of students logging unreliable browsing per minute during the lecture. The highest average is approximately 6 students, and this value is 3.6% of all students in the lecture. Overall, the browsing behavior of approximately 98% of students was grasped by real-time processing on an average.

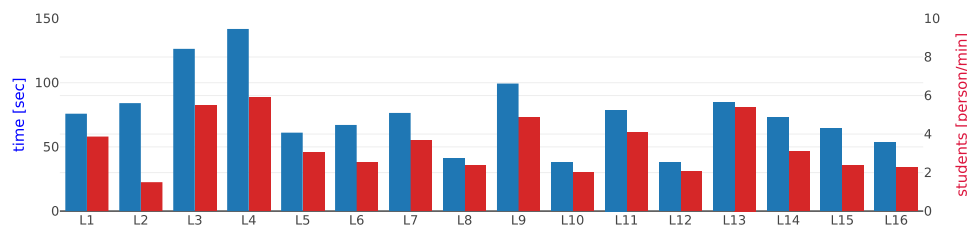


Figure 1: *Unreliable browsing time. Blue bars depict average unreliable browsing time of each student in each lecture. Red bars depict average number of students logging unreliable browsing per minute during the lecture.*

Table 1 shows the average missing browsing time for one student per minute. The shorter the process cycle, the shorter missing browsing time calculated. More specifically, when the processing interval t is 300 seconds, approximately 20 seconds of browsing time is ignored per minute; further, when the processing interval is 10 seconds, less than 3 seconds of browsing time is ignored on an average. This result indicates that shortening the processing interval leads to more accurate representation of information obtained from the event log.

Finally, the observations for the processing time of each processing interval t is provided in Table 2. We measured the time when browsing behavior was calculated and then recorded into a database. In this experiment, we used a server consisting of a XeonREG E5-2667 of 3.2GHz CPU and 256GB memory. If there were approximately 150 students in a lecture, the processing time is sufficiently rapid.

Table 1: *Missing browsing time* [sec / min]

t [sec]	ave.	std.
10	2.79	1.71
20	4.27	2.20
30	5.54	2.63
60	8.65	3.74
120	13.28	5.25
300	21.48	7.75

Table 2: *Processing time* [sec]

t [sec]	ave.	std.	max
10	1.10	0.28	2.94
20	1.24	0.36	2.88
30	1.36	0.42	3.15
60	1.61	0.51	3.26
120	1.86	0.60	4.00
300	2.21	0.78	5.24

4 CONCLUSION AND FUTURE WORK

In this paper, we evaluated the validity of a tracking system for page-browsing in our real-time feedback system (Shimada 2018a) with regards to two factors: reliability of tracking browsing behavior and an appropriate processing cycle for real-time feedback. In conclusion, our system demonstrated enough reliability, and considering the necessity to process multiple lectures simultaneously at universities and the trade-off between missing browsing time and processing time, the processing interval of 30 seconds seems effective for the real-time system. In a future study, we will conduct further investigations and present guidelines for using a real-time system.

ACKNOWLEDGEMENTS

This work was supported by JST AIP Grant Number JPMJCR19U1, and JSPS KAKENHI Grand Number JP18H04125, Japan.

REFERENCES

- A. Shimada, S. Konomi, H. Ogata (2018a). Real-time learning analytics system for improvement of on-site lectures. *Interactive Technology and Smart Education*, 15(4), 314-331.
- A. Shimada, Y. Taniguchi, F. Okubo, S. Konomi, H. Ogata (2018b). Online change detection for monitoring individual student behavior via clickstream data on E-book system. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* pp. 446-450.
- H. Ogata, C. Yin, M. Oi, F. Okubo, A. Shimada, K. Kojima, M. Yamada (2015). E-Book-based learning analytics in university education. *Proceedings of the 23rd International Conference on Computers in Education (ICCE 2015)* pp.401-406.

LAK Theory 2020: Workshop on Theory and Learning Analytics

Kathryn Bartimote

University of Sydney

kathryn.bartimote@sydney.edu.au

Sarah K. Howard

University of Wollongong

sahoward@uow.edu.au

Dragan Gasevic

Monash University

dragan.gasevic@monash.edu

ABSTRACT: The workshop addresses the ongoing question of connecting theory and learning analytics.

A key aim of the workshop is to address issues identified in the LAK 2019 Educational Theory session, specifically developing further clarity around concept definition, and the role of theory in design, model validation and interpretation of findings. The organisers will set the scene by giving an overview of theory use in learning analytics, with a particular emphasis on sociological and psychological theories and their application. Participants will be invited to nominate a current research project that would benefit from a roundtable-style discussion with colleagues, along with a theoretical framework of interest. Expected outcomes are the formation of a community of practice, a publication on challenges of and opportunities for theory use in the field, and a template for an ongoing workshop initiative. To support the community, an online space will be created for pre-workshop preparation and ongoing collaboration.

Keywords: theory, research cycle, design, model validation, interpretation

1 BACKGROUND

This workshop is founded on the premise that the quality of learning analytics, both research and practice, rests on the strength of its connection to theory (Gašević, Dawson, & Siemens, 2015). Through this workshop we hope to build an ongoing community of scholars interested in both using educational (and other) theory in learning analytics research and practice, and contributing to further development of theory through their work.

Theory provides a common language through which to communicate about research, it gives a frame of reference to understand the type of knowledge being generated, and what may be legitimately claimed (Reimann, 2016). In a typical research cycle, we suppose that theory influences the questions we ask, design of data collection, analysis approach and method, and interpretation and reporting of results (Wise & Shaffer, 2015). In this way we are arguing for a move away from the primacy of method

in learning analytics, that is, away from pragmatism to theory-driven paradigms for research where theory underpins method and the two cannot be separated (Bartimote, Pardo, Reimann, 2019). This adds the possibility for explanation – for an observed pattern, for a prediction, for why an intervention or pedagogical strategy works – in research, and in practice.

Theory allows for informed practice by a range of actors that support learning in educational settings, such as teachers, student support officers, advisors, and academic managers. If the objective of learning analytics is actionable information, then theory-driven analytics enables choices and decisions that are situated in defensible frameworks (Bartimote, Pardo, Reimann, 2019). And it means we have a starting point for explanation when things do or don't work, and a basis for adaption of tactics and strategies shown to be effective in one context, in other contexts. For analysts, data scientists, and software developers, theory may guide what activities to capture, the development of indicators and measures, the display of information, and the form of personalised messages and automated nudges. We need to focus on providing information about constructs that matter, and learning (and other) theories substantiated by empirical research can serve as useful starting points.

The LAK community is increasingly drawing on ideas from the learning sciences, educational psychology, sociology, and social psychology. This is demonstrated by the inclusion of an educational theory session on the LAK 2019 program, but more generally in recently published learning analytics work referring to theories such as social cognitive theory and self-efficacy beliefs, various self-regulated learning models, measurement theory, social-constructivism, human-computer interaction (HCI) and activity theory, Kolb's experiential learning cycle, etc. We consider the time is ripe for a call across the community to gather to consider more explicitly the role of theory in learning analytics.

Moving on from the LAK 2019 Educational Theory session, it is necessary to address some of the issues raised regarding definitions of concepts, design, model validation and interpretation of findings. To do this, multidisciplinary groups of researchers working in the area need to come together to support this work and begin to create some level of understanding in the field. This is the work proposed for the LAK 2020 theory workshop.

2 ORGANISATIONAL DETAILS

2.1 Proposed Half-Day Workshop Schedule

Table 1: Proposed schedule.

Timing	Description	Contributors
10 minutes	Welcome and setting the scene	Organisers
30 minutes	Introductory plenary 'Overview of theory use in learning analytics': 20 minutes presentation, followed by 10 minutes Q&A	Organisers
50 minutes	Soapbox session ¹ : 5 minutes to espouse your favourite theory and argue its relevance to learning analytics, then 5 minutes audience rebuttal and critique	Participants: 5 presenters

¹ Soapbox session presenters will be required to submit a 250 word abstract.

30 minutes	Morning tea	
60 minutes	Work in progress roundtables ² : 10 minutes to introduce project, summarise progress to date, outline challenges to be overcome, and input that would be useful from the group, followed by 10 minutes discussion with colleagues at table	Participants: 3 research teams per roundtable group
20 minutes	Roundtable report back: group representatives to summarise conversation and potential impact on the work	Participants
10 minutes	Next steps plenary discussion, and close: Gauge interest in further activities around theory and learning analytics e.g. LAK 2021 2nd annual workshop, publications, LASI 2021 workshop/tutorial	Organisers

2.2 Other details

The event will be an open workshop. All attendees will have the opportunity to give a short presentation on either a theory and/or work in progress, should they wish to, as detailed in the schedule above. Abstract submissions of 250 words for these short presentations will be handled via the event's Google Site: <https://sites.google.com/view/lak20theoryworkshop/home>. The submission timeline will follow the timeline suggested by the conference organisers, that is, call for participation 29 October 2019, and notification of acceptance in time for early-bird registration deadline 20 January 2020. We anticipate a registration of 15-20 participants. Please use #LAKtheory when referencing this event on social media.

3 OBJECTIVES/INTENDED OUTCOMES

The workshop will provide a space for both capacity building and connection, and it is hoped that the event will spark the formation of a community of practice. The outcomes of the event will be housed on the Google Site. The possibility of a single publication or edited collection will be discussed amongst organisers and participants, and this event will serve as a template for an ongoing workshop initiative on theory and learning analytics.

4 WEBSITE STRUCTURE AND CONTENT

The Google website will: 1. support pre-workshop data gathering and planning materials; 2. act as a collection point for materials, group interactions and archive for the workshop; and, 3. support ongoing dissemination and group activities. It is the aim that the workshop is ongoing, in which case the website will be an ongoing hub for year to year activities and building field memory. The structure of the website is based on theory informing the research cycle, at three stages: design, method, interpretation. Each of these stages will be a section of the website. The website will include: About, Back-

² Roundtable session presenters will be asked to indicate the stage of their work at the time of submission of a 250 word abstract e.g. data collection/extraction, data analysis, write up. Where possible, presenters at a similar stage will be grouped together.

ground literature, Workshop materials, Working areas: Design, Method, Interpretation. In the workshop, participants will be grouped based on where they self-identify their work within the three stages, at the time of attendance. Over time, as work develops and builds, they may access resources to support ongoing development.

REFERENCES

- Bartimote, K., Pardo, A., & Reimann, P. (2019). The perspective realism brings to learning analytics in the classroom. In J. M. Lodge, J. Cooney Horvath, & L. Corrin (Eds.), *Learning Analytics in the Classroom: Translating Learning Analytics Research for Teachers*. London: Routledge.
- Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64-71.
- Reimann, P. (2016). Connecting learning analytics with learning research: The role of design-based research. *Learning: Research and Practice*, 2(2), 130-142.
- Wise, A. F., & Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics*, 2(2), 5-13.

Let's Talk LA: Discussing Challenges for Institutional Adoption of Learning Analytics

Isabel Hilliger

Pontificia Universidad Católica de Chile
ihillige@ing.puc.cl

Yi-Shan Tsai

University of Edinburgh
Yi-Shan.Tsai@ed.ac.uk

Pedro J. Muñoz-Merino

Universidad Carlos III de Madrid
pedmume@it.uc3m.es

Mar Pérez-Sanagustín

Université Paul Sabatier Toulouse III
mar.perez-sanagustin@irit.fr

ABSTRACT: Regardless the proliferation of Learning Analytics (LA) initiatives over the past years, their adoption in Higher Education (HE) institutions remains immature. In this interactive workshop, we invite researchers, practitioners and policy makers to discuss the main challenges for the adoption of LA initiatives at institutional level. Using a set of instruments developed as part of the European Erasmus + Project LALA, participants will engage with the current state of LA in different HE institutions, the main challenges they are facing, and key actions taken or to be taken to address them. The inputs from the participants will be turned into a manifesto for the institutional adoption of LA in HE.

Keywords: learning analytics adoption, higher education, stakeholder engagement

1 MOTIVATION AND BACKGROUND

Regardless of all the frameworks and instruments that have been proposed over the past decade for starting Learning Analytics (LA) initiatives (see Table 1), LA adoption remains immature (Dawson et al., 2018; Tsai, Poquet, Gašević, Dawson, & Pardo, 2019; Viberg, Hatakka, Bälter, & Mavroudi, 2018). Not only are LA tools not used effectively in shaping decision-making among teaching staff (Klein, Lester, Rangwala, & Johri, 2019), but there is also insufficient evidence of improved student learning outcomes through LA tool usage (Klein et al., 2019; Viberg et al., 2018). On the one hand, LA research is still in its infancy, and there is a need for more studies that focus on evaluating the impact of LA and the progress of the field as a whole (Viberg et al., 2018). On the other hand, LA researchers and Higher Education (HE) leaders continue to experience challenges, such as the lack of stakeholder buy-in (Tsai et al., 2019), resulting in LA initiatives not being scaled or adopted in a systematic manner.

It is important to bring different stakeholders together to discussing how to scale up LA initiatives, allowing HE researchers, practitioners and policy-makers to exchange ideas about how LA tools could be use in everyday decision-making (Ferguson et al., 2016). So far, there has been positive progress throughout the collaboration of different universities in multi-national projects funded by the European Commission, such as the SHEILA (<https://sheilaproject.eu/>) and the LALA projects (<https://www.lalaproject.org>). These projects propose guidelines and frameworks to help HE institutions installing and adopting LA initiatives at a large scale.

Table 1: Frameworks and instruments published over the past decade to provide institutions with guidelines for Learning Analytics adoption

<i>Framework or instrument</i>	<i>Purpose</i>	<i>References</i>
Learning Analytics Framework	Determine critical dimensions for setting up LA services	(Greller & Drachsler, 2012)
Maturity Index	Measure the institutional progress in LA Adoption	(Bichsel, 2012)
Learning Analytics Sophistication Model	Illustrate the different stages of LA implementation	(Siemens, Dawson, & Lynch, 2013)
Learning Analytics Readiness Instrument (LARI)	Help institutions to prepare themselves for LA adoption	(Arnold, Lonn, & Pistilli, 2014)
ROMA Outcome Mapping Approach	Provide institutions with practical steps for LA policies	(Ferguson et al., 2014; Macfadyen, Dawson, Pardo, & Gasevic, 2014)
Evaluation Framework for Learning analytics (EFLA)	Measure and compare the impact of LA initiatives	(Scheffel, 2017)
SHEILA Framework	Inform strategic planning and policies for LA adoption	(Yi-shan Tsai et al., 2018)

This workshop builds upon the prior experience of these two projects, using a participatory approach to understand the current state of LA adoption in different HE settings and identify common challenges. Specifically, we will use two instruments that have been developed based on the experience of the four Latin American partners of the LALA project. First, we will use an instrument called LALA Canvas, which is a template based on the ROMA Outcome Mapping Approach for guiding group discussions about the current state of LA adoption in HE institutions (<http://bit.ly/LALACanvasEn>). Second, we will use the LALA Map, which is a graphical representation of the current state of LA adoption in different HE institutions in terms of maturity levels and the types of leadership present to drive the use of LA tools (see Figure 1). Both instruments have already been used to facilitate conversation on LA adoption and maturity among LA researchers and HE practitioners from diverse universities in Europe and Latin America.

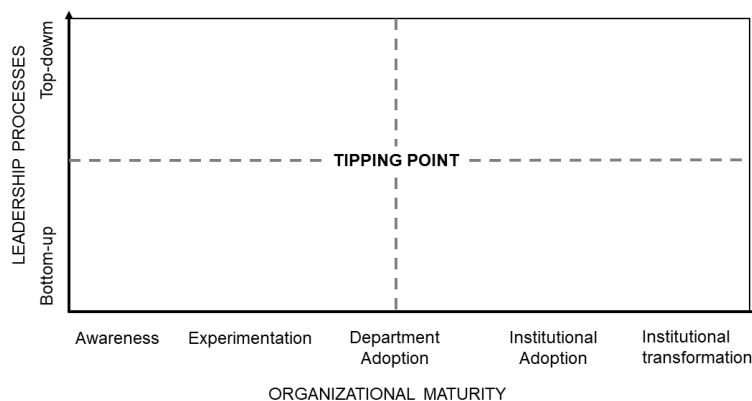


Figure 1: LALA Map for understanding the current state of LA adoption in different institutions

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

2 WORKSHOP OBJECTIVES

The general objective of this workshop is to engage LA researchers, policy makers and HE practitioners to discuss the current state of LA adoption, in addition to providing them with instruments for exchanging ideas about how to scale LA adoption in their institutions:

- Presenting the current state of the art on frameworks and instruments for LA adoption.
- Introducing the LALA Canvas and the LALA Map as instruments for guiding discussions about the current state of LA adoption in their institutions.
- Identifying commonalities and differences among HE institutions in terms of LA adoption by using the LALA Canvas and the LALA Map throughout the workshop.
- Develop a *manifesto* about what workshop participants expect to achieve in terms of institutional adoption of LA.

3 WORKSHOP FORMAT AND ORGANIZATIONAL DETAILS

This **open workshop** will consist of a **half-day participatory session** (between 1:30 pm and 5 pm), open to any researcher, policy maker and practitioner motivated to scale LA adoption (**expected number of 25 participants**). To facilitate knowledge exchange, we will combine **presentations** and **participatory activities**, using the LALA Canvas and the LALA map as tools to facilitate group discussions (see more **organizational details** in the following link: <http://bit.ly/LAK20WSDetails>). The LALA Canvas has already been applied in four workshops, having participants from Latin America, North America and Europe. As for the LALA Map, it emerged from the LA adoption experience (designing and implementing LA tools) of the four Latin American partners of the LALA project. However, the LALA Map is still applicable to institutions outside the Latin American context, as the key dimensions were developed based on LA literature. The maturity level is based on the interpretation of prior work conducted by Bichsel (2012) and Siemens et al. (2013), and LA leadership is presented as a spectrum between top-down and bottom-up approaches described by Dawson et al. (2018). The workshop will proceed in the following steps:

1. The organisers will present existing challenges in LA adoption, in addition to a brief review of existing frameworks and instruments for starting and scaling LA initiatives.
2. The organisers will introduce the LALA Canvas and its dimensions. The workshop participants will work in small groups to analyse each of the LALA Canvas' dimensions, listing elements for each dimension according to the current context of their institutions.
3. The moderators will introduce the LALA Map, including an explanation of how this map describes the current state of LA adoption in the four Latin American partners affiliated to the LALA project. Workshop participants will then be invited to locate their institutions in the LALA map, and discuss the commonalities and differences in LA adoption among different HE institutions.
4. The inputs from the participants will be turned into a *manifesto*, which will outline what they expect to achieve in their institutions in terms of LA adoption

4 PROGRAM COMMITTEE AND EXPECTED PARTICIPANTS

The program committee will be formed by the four authors. Additionally, other researchers of the LALA Project will be involved during the workshop to support discussions within participants. Regarding expected participants, this is an open workshop for HE practitioners, policy makers, and LA researchers who are interested in adopting LA systematically and effectively.

ACKNOWLEDGEMENTS

This work is funded by the EU Commission under the LALA project (grant no. 586120-EPP-1-2017-1-ES-EPPKA2-CBHE-JP). This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained.

REFERENCES

- Arnold, K. E., Lonn, S., & Pistilli, M. D. (2014). An Exercise in Institutional Reflection: The Learning Analytics Readiness Instrument (LARI). In *LAK'14: International Conference on Learning Analytics and Knowledge*. Indianapolis, IN, USA.
- Bichsel, J. (2012). *Research Computing: The Enabling Role of Information Technology*. Louisville, CO: ECAR. Retrieved from <http://www.educause.edu/ecar>
- Dawson, S., Poquet, O., Colvin, C., Rogers, T., Pardo, A., & Gasevic, D. (2018). Rethinking learning analytics adoption through complexity leadership theory. In *LAK'18: International Conference on Learning Analytics and Knowledge*. Sydney, NSW, Australia.
- Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., ... Vuorikari, R. (2016). Research Evidence on the Use of Learning Analytics: Implications for Education Policy. In V. R. & J. Castaño Munoz (Eds.), *Joint Research Centre Science for Policy Report* (pp. 1–150). Luxembourg: Publications Office of the European Union. <https://doi.org/10.2791/955210>
- Ferguson, R., Macfadyen, L. P., Clow, D., Tynan, B., Alexander, S., & Dawson, S. (2014). Setting Learning Analytics in Context: Overcoming the Barriers to Large-Scale Adoption, 1, 120–144.
- Greller, W., & Drachsler, H. (2012). Translating Learning into Numbers: A Generic Framework for Learning Analytics, 15, 42–57.
- Macfadyen, L. P., Dawson, S., Pardo, A., & Gasevic, D. (2014). Embracing big data in complex educational systems: The learning analytics imperative and the policy challenge. *Research & Practice in Assessment*, 9(2), 17–28. <https://doi.org/10.1017/CBO9781107415324.004>
- Scheffel, M. (2017). *The Evaluation Framework for Learning Analytics Impact The Evaluation Framework*. The research reported in this thesis was carried out at the Open Universiteit in the Netherlands at Welten Institute – Research Centre for Learning, Teaching and Technology, and under the auspices of SIKS, the Dutch Research School for Information and Kno.
- Siemens, G., Dawson, S., & Lynch, G. (2013). *Improving the Quality and Productivity of the Higher Education Sector Policy and Strategy for Systems-Level Deployment of Learning Analytics*. Sydney, Australia: Australian Government Office for Learning and Teaching.
- Tsai, Yi-shan, Gasevic, D., Whitelock-Wainwright, A., Muñoz-Merino, P. J., Moreno-Marcos, P. M., Rubio Fernández, A., ... Kollom, K. (2018). *SHEILA Supporting Higher Education to Integrate Learning Analytics*.
- Tsai, Yi-Shan, Poquet, O., Gašević, D., Dawson, S., & Pardo, A. (2019). Complexity leadership in learning analytics: Drivers, challenges and opportunities. *British Journal of Educational Technology*, 0(0), 1–16. <https://doi.org/10.1111/bjet.12846>
- Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior*, 89(July), 98–110. <https://doi.org/10.1016/j.chb.2018.07.027>

Bringing together writing tool design, writing analytics and writing pedagogy

Christian Rapp¹, Susan Lang², Antonette Shibani³, Kalliopi Benetos⁴, Chris Anson⁵

¹ZHAW, School of Management & Law, Center for Innovative Teaching & Learning, Winterthur, Switzerland; ²Center for the Study & Teaching of Writing, Ohio State University, Columbus, OH, USA; ³Faculty of Transdisciplinary Innovation, University of Technology Sydney, Australia; ⁴TECFA Educational Technologies Unit, University of Geneva, Switzerland;

⁵Campus Writing & Speaking Program, North Carolina State University, Raleigh, NC, USA.

¹rapp@zhaw.ch, ²lang.543@osu.edu, ³antonette.shibani@uts.edu.au,

⁴kalliopi.benetos@unige.ch, ⁵chris_anson@ncsu.edu

ABSTRACT: The evolution of digital technologies and the writing tools that have subsequently been developed from them opened the way for the emergence of writing analytics as a field of academic research. Within digital writing tools, writing analytics are used to gather and analyze data for research, and to provide automated feedback for writers and insights for instructors. Writing analytics methods and tools can help improve our understanding of writing processes and products. Current reviews of digital writing tools show that much of what writing analytics has to offer has been garnered for the purposes of automating evaluation and scoring, leaving an application gap for writing tools that support pedagogies aiming to develop effective writing strategies. Building upon the development of writing analytics methods and tools can help future tool designs to better support effective writing pedagogy and practice, and suggest future foci for writing analytics advancement. This proposed workshop aims to bring together writing pedagogy researchers, writing instructors, writing tool developers, and writing analytics specialists in order to explore the potential contributions of their respective fields in the development of effective digital writing environments, and also to provide a forum for the planning of future collaborative works.

Keywords: writing analytics, learning analytics, collaborative writing, writing theories, writing tool development

1 BACKGROUND

Recent years have seen a mushrooming of digital tools supporting writing and its instruction, with new additions appearing at an increasing pace. A review of computer-based writing instruction identified automated essay scoring and automated essay evaluation systems that assess and provide feedback on student essays, with progress made toward adaptive and personalized writing tools such as Intelligent Tutoring Systems (ITS) (Allen, Jacovina, & McNamara, 2015). Recently, Strobl, et al. (2019) identified 89 academic writing tools supporting writing in secondary and higher education. One finding related to this workshop is existing classifications failing to grasp not only the increasing breadth of functionality, but also overlooking pedagogies and practices within which they are being used. One way to better understand tool development in the context of writing pedagogies is through writing analytics. While still an emerging field of research, we suggest that writing analytics, in a broader definition, can support writing tool development and vice versa to the mutual benefit of both areas.

“Writing analytics” was defined by Buckingham Shum et al. (2016) as “involv[ing] the measurement and analysis of written texts for the purpose of understanding writing processes and products, in their educational contexts. Writing analytics are ultimately aimed at improving the educational contexts in which writing is most prominent” (p. 481). This emerging field “equally invokes methodological processes and the theory and content of writing instruction” as it applies a variety of data-driven lenses to writing instruction processes and products (Lang, Aull, & Marcellino, in press). In doing so, writing analytics supports the ongoing development of various writing tools, both through analysis of artifacts produced using such tools, and in guiding the development of tools that focus on assessment and measurement of individual and aggregated data. Writing analytics projects can examine the features of tools and the artifacts produced through additional features. Writing analytics significantly extends traditional human computer Interaction writing tool analysis. Writing analytics and data from writing tool usage can furthermore be visualized and fed back to learners, instructors, tool developers, and researchers (Rapp & Ott, 2017; Vieira, Parsons, & Byrd, 2018).

In learning analytics, text features have been studied using linguistic tools to understand language better. Tools like Coh-Metrix and WAT identified indices of text based on cohesion, language, complexity, and readability, which were used to study various writing dimensions (Crossley, Allen, Snow, & McNamara, 2015). In addition, writing processes like drafting and revision are studied using fine-grained data from the trace logs from individual and collaborative writing settings (e.g., Shibani, Knight, & Buckingham Shum, 2018). Writing analytics tools providing automated feedback have much improved, e.g., with contextualizing feedback for disciplinary contexts (in AcaWriter) by co-designing tool and instructor feedback, and then integrating within curricula (Shibani, Knight, & Buckingham Shum, 2019). Along with research tools, proprietary software providing automated feedback on writing, e.g., Revision Assistant by Turnitin (Woods, Adamson, Miel, & Mayfield, 2017) and Writing Mentor Google add-on by ETS (Madnani et al., 2018), advances writing tool capabilities. Consequently, tools need to align to established pedagogy for effective usage, referring to writing instruction studies by incorporating writing pedagogies within writing analytics (Graham & Perin, 2007).

2 WORKSHOP FOCUS

This workshop was inspired by the 8th International Conference on Writing Analytics, Winterthur, Switzerland (<https://writinganalytics.zhaw.ch/>). Diverse tools developed by European and North American scholars were presented, with most implemented as a Software-as-a-Service, and therefore collected large amounts of usage data. While tools concentrate on writing data collection (e.g., keylogging by Inputlog) or automated feedback provision for written text (e.g., Writing Aid Dutch/Academic Writing Assistant), others (e.g., Thesis Writer, Research Writing Tutor, C-SAW, AcaWriter) combined facilitated system support logging (e.g., tutorials or phrasebooks) with text production and revision, allowing inquiry into the uses and effects of support functions on subsequent text production (or revisions). However, it is far from clear what data should be collected, how it should be analyzed (and potentially displayed), for what purpose, and for what audiences.

The aim of this proposed LAK Writing Analytics Workshop is to draw upon the results of previously held meetings, as well as the most recent research, and to bring together writing tool developers,

writing analytics specialists, and writing pedagogy instructors and researchers in order to discuss (1) current practices in their respective fields, (2) opportunities, research questions and corresponding designs for collaborative works, and (3) resulting design choices for future tool development and/or writing analytics research agendas that could be informed by current and prospective developments.

The workshop will focus on the following questions: (1) Which tools collect data suitable for writing analytics? What data is collected and for what purposes, how it is analyzed, and for whom? (2) Which writing analytics methods are currently being employed, for which audiences, and for what purposes? (3) Concerning the linkage between theory and practice, writing theorists (e.g., Graham & Perin, 2007) have proposed ways to foster the learning of academic writing. In what ways can Writing Analytics support this, and what are the implications for writing tool developers? (4) How can we create better synergies among writing tool developers, writing analytics specialists and practitioners, and writing pedagogy researchers? Are there any lessons for writing analytics common to secondary and higher educational writing contexts, and that are also appropriate across different geographical contexts?

3 SUBMISSIONS AND WORKSHOP FORMAT

Workshop activities and schedule

To achieve the goals of this half-day workshop as articulated, we propose the following design:

Welcome: Short introduction(s) of participants (5 minutes. 0900-0905)

Overview: Short overview of the field (15 minutes. 0905-0920)

Input-phase: Short statements of accepted papers along three lines: (60 minutes. 0920-1020)

1. Perspective – writing tool developers/ users.
2. Perspective – writing practice/ pedagogy.
3. Perspective – writing analytics.

Working phase: Discussion within the three groups along the suggested following questions (What are good current practices? Where do we want to be in 2-3 years? What do we need from the other groups to get there?) (60 minutes. 1030-1100)

Results: Presentations from all three groups (10 minutes per group, 30 minutes in total. 1100-1130)

Discussion/ synthesis: All three groups to take part in a discussion plenum (30 minutes. 1130-1200)

Future steps: Discussion of the future developments of WA, its application, and the wider WA community. (30 minutes. 1200-1230)

Participation and Dissemination

The workshop will be of interest to a wide range of LAK delegates including students and researchers engaged in writing research and the use of writing tools; educators in schools, universities and businesses; data analysts; and companies active or potentially active in the field. An open call will be made for submissions via a website. Workshop organizers will make use of listservs and their own personal networks to advertise the workshop. The European location of LAK20 provides an opportunity to strengthen the European community in writing analytics and writing tool developments and to link with colleagues across continents. The workshop was announced at the 8th International Writing Analytics Conference (Winterthur Switzerland) and will be held prior to the next European Writing Analytics Conference (Fall, 2020). At least one board member of the European

Association for Teaching Academic Writing will participate. Selected participants will be invited to work with the editors of The Journal of Writing Analytics in order to propose and develop brief manuscripts for publication in Volume 4.

REFERENCES

- Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2015). Computer-based writing instruction. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 316-329). New York, NY: Guilford Press.
- Buckingham Shum, S., Knight, S., McNamara, D., Allen, L., Bektik, D., & Crossley, S. (2016). Critical perspectives on writing analytics. In *Proceedings of the sixth international conference on learning analytics & knowledge (LAK'16)* (pp. 481-483). ACM.
- Crossley, S., Allen, L. K., Snow, E. L., & McNamara, D. S. (2015). Pssst... textual features... there is more to automatic essay scoring than just you! In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge 2015* (pp. 203-207). New York, NY: ACM.
- Graham, S., & Perin, D. (2007). A Meta-Analysis of Writing Instruction for Adolescent Students. *Journal of Educational Psychology*, 99(3), 445-476.
- Lang, S., Aull, L., & Marcellino, B. (in press). Constructing a Foundational Taxonomy of Writing Analytics. *Journal of Writing Analytics*.
- Madnani, N., Burstein, J., Elliot, N., Klebanov, B. B., Napolitano, D., Andreyev, S., & Schwartz, M. (2018). Writing Mentor: Self-Regulated Writing Feedback for Struggling Writers. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations* (pp. 113-117). Stroudsburg, PA: Association for Computational Linguistics.
- Rapp, C., & Ott, J. (2017). Learning Analytics in Academic Writing Instruction—Opportunities Provided by Thesis Writer (TW). In C. Igel, C. Ullrich, & M. Wessner (Eds.), *Bildungsräume Proceedings 2017* (pp. 391-392). Bonn, Germany: Gesellschaft für Informatik.
- Shibani, A., Knight, S., & Buckingham Shum, S. (2018). Understanding Revisions in Student Writing through Revision Graphs. In C. Penstein, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta,...B. du Boulay (Eds.), *Artificial Intelligence in Education 19th International Conference, AIED 2018* (pp. 332-336). Springer.
- Shibani, A., Knight, S., & Buckingham Shum, S. (2019). Contextualizable Learning Analytics Design: A Generic Model, and Writing Analytics Evaluations. In D. Azcona & R. Chung (Eds.), *Proceedings of the 9th International Conference on Learning Analytics and Knowledge (LAK'19)* (pp. 210-219). New York, NY: ACM.
- Strobl, C., Ailhaud, E., Benetos, K., Devitt, A., Kruse, O., Proske, A., & Rapp, C. (2019). Digital Support for Academic Writing: A Review of Technologies and Pedagogies. *Computers & Education*, 131, 33-48.
- Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education*, 122, 119-135.
- Woods, B., Adamson, D., Miel, S., & Mayfield, E. (2017). Formative Essay Feedback Using Predictive Scoring Models. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 2071-2080). New York, NY: ACM.

A Framework for Assessing Reflective Writing Produced Within the Context of Computer Science Education

Huda Alrashidi^{1,4}, Thomas Daniel Ullmann², Samiah Ghounaim³, Mike Joy¹

Computer Science Department, University of Warwick¹

Institution Institute of Educational Technology, The Open University²

Centre for Applied Linguistics, University of Warwick³

h.alrashidi@warwick.ac.uk⁴

ABSTRACT: Reflective writing is known to be an effective activity to increase students' learning. However, there is limited literature in reflective writing assessment criteria in the context of computer science (CS) education. In this paper, we aim to explore a meaningful reflective writing assessment characteristics. That has been used to assess reflective text by CS educators. This paper has two contributions: (a) we developed a Reflective Writing Framework (RWF) for the main criteria has been used to assess reflective text in CS education from the findings of a semi-structure questionnaire; (b) the RWF was tested empirically using a pilot test of the manual annotation used to modify the framework. This analysis resulted in an inter-rater reliability of 0.78 being achieved. The overall goal of this research is to develop a Learning Analytics (LA) tool which can automatically detect the categories of the RWF present in a text to assess the student authors' reflective writing in relation to CS.

Keywords: Reflective Writing, Computer Science, Reflection, Reflection Detection, Reflective Writing Analytics, Learning Analytics

1 INTRODUCTION

Learning Analytics (LA) is gradually becoming one of the pivotal aspects of educational technology. This paper investigates an LA tool that supports reflection by analyzing and providing feedback on reflective writing (RW). RW can support students to gain awareness of their learning processes. In terms of Computer Science (CS), "reflection is worth encouraging, for its indirect effect on the technical skills and knowledge which are our ultimate purpose in teaching computer science" (Fekete, Kay, Kingston, & Wimalaratne, 2000). Technical skills are at the core of CS, and these center around formulating problems and their solutions. Since reflection is a metacognitive process, it can only be assessed indirectly - through written or verbal forms. Analyzing RW manually makes giving students feedback a challenging and time-consuming task. Automated feedback can better support the students in terms of providing timely analyses. LA tools have the goal of supporting reflection – specifically, by analyzing students' reflective texts. To design an LA tool for RW, there is a necessity either to adapt an existing methodology or to develop a new framework for this purpose (Gibson et al., 2017). This study aims to develop a RW Framework (RWF) for CS education to develop an LA tool for RW. We focus on the following research questions: 1) what are the characteristics of RW within CS education? And 2) what are the indicators which can be used to assess RW levels as they occur in CS education?

2 RELATED WORK

Reflective activities that have been used recently investigated RW in CS education (Alrashidi, Joy, & Ullmann, 2019; George, 2002; Stone & Madigan, 2007) as they have in other disciplines such as social and health sciences. However, the literature on RW in CS education is limited. For instance, George (2002) and Fekete et al. (2000) investigated using the reflective journal in terms of benefits to the students in an undergraduate programming course. Both studies noted that reflective journals were beneficial to get students to reflect on their software development processes as it is part of their learning outcome. Moreover, in accordance with the LA tool for reflection in CS education, Dorodchi et al. (2018) implemented an activity based on the CS course with periodic reflection by applying Kolb's learning model. They validated the result of student reflection through the LA classification model. It concluded that including reflection as a feature could improve the accuracy and time of their classification model. However, there is a difficulty for students to reflect effectively on their own understanding. Moskal and Wass (2019) developed an approach for educators to encourage students to think about their software development steps through a series of sessions. They found that the approach was beneficial for both students and educators. However, Grossman (2009) mentioned that a number of students did not understand what they are expected to reflect on due to lack of guidance. Grossman's (2009) findings provide reasoning for the study conducted by George (2002) that found reflective journal as not widely accepted by students and/or educators in CS education. The RWF developed here is a guideline for students to determine the main elements on which they are expected to reflect, and for educators on assessing their students' RW.

3 THE RWF

Semi-structured questionnaires explored perspectives of 6 HE experts (Exp.) –selected based on their breadth of academic skills in CS and their knowledge of reflection– on RW levels, and the indicators they use to assess RW in CS education. A thematic analysis of the responses resulted in three codes for levels of reflection: 1) non-reflective, 2) reflective, and 3) critically reflective; and seven codes for indicators summarized as follows.

First, the **descriptive**: two experts used similar words in their definitions of such indicators. Exp. A stated that: “students merely **describe** what they have done ... without any examples.” Exp. C used the word “**listing**” instead stating that “I would often see simple summaries of lesson content, or **listings** of topics covered that I would class as **non-reflective**”. This means that “**non-reflective**” texts are superficial descriptions of situations. Second, the **understanding**: all the experts characterized it as bordering on the reflective level. For example, Exp. E defined this indicator as, “when students **identify** their **understanding** of competencies ... [RW] has been reached.” Accordingly, the understanding indicator characterizes both the non-reflective and the reflective levels, per the context. Third, the **feeling**: the experts argued that the reflective level applies when the writer can identify their own thoughts and feelings. For example, Exp. C stated that “I would look for evidence of what the students previously **thought** or **felt** on whether that had worked or not.” This means that the feeling indicator in the proposed framework can be either at the reflective or critically reflective level. Fourth, **reasoning**: they argued it occurs when a writer explains a situation/issue by providing examples/causes. For instance, students would “clearly **explain** their process, what worked, what didn't” (Exp. D), and “**provide examples**” (Exp. G), and/or “**analysis** of problems and [their solutions]” (Exp. C). Fifth, **perspective**: this could be detected when “Students share **personal** thoughts and

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

connect with **other** thoughts” (Exp. G), and giving “evidence of **re-evaluation** [due to] feedback from **others**” (Exp. D). Both experts emphasized that perspective takes into consideration others’ perspectives. Exp. D summarizes it as students’ ability “to **connect** the topic in question to wider applications in the **discipline**, their **community**, or the **world**”.

Sixth, the significance of the **new learning indicator** was clearly emphasized by the panels. The experts commented that they search for evidence of learning. For example, Exp. H mentioned that the student must show evidence of what has been learnt in terms of personal and professional skills by “connecting what we have learned and the skills ... gained to our own personal or professional developments”. Lastly, **future action**: the panel of experts commented that they search for evidence of outcome when assessing RW. Exp. C expected the student to show they had achieved a deeper understanding of the problem they were engaged with, as a result of producing the RW, in terms of *cognition* by having “ a **deeper** understanding of what they have learnt”, *metacognition* by being “**better** able to manage their **own learning** and **development** once they leave formal education,” and *socially* with the ability “to work **better** in a team by identifying and owning their own weaknesses, and sharing their successes.”

Table 1 shows all the indicators and levels of our RWF which is consistent with the literature on RW and on reflection theories, especially in terms of the levels defined by Wong, Kember, Chung, and Yan (1995) and the reflection indicators defined by Ullmann (2019).

Table 1 Levels and Indicators of the RWF for CS

Reflective levels	Indicators
Non- Reflective	Descriptive: the writer reports a fact from experience and/or materials
	Understanding: the writer understands and/or analyses the experience.
Reflective	Feelings: the writer identifies and/or analyses their own thoughts and feelings.
	Reasoning: the writer explains the experience by giving reasons.
Critically	Perspective: the writer shows awareness of alternatives.
Reflective	New learning: the writer integrates and/or describes new learning
	Future action: the writer intends and/or plans to do something in the future.

4 VALIDATION OF THE RWF

A manual annotation aimed to produce a final version of the framework through manual reviews and using this activity as a basis for an iterative cycle of framework development. The dataset consisted of 30 RW documents –split into 360 sentences– from 30 computer science students in module CS310 Computer Science Project all relating to a 3rd-year project undertaken during 2013–2016 academic years. The data were collected by the CS Department at the authors’ university as part of its normal assessment process and then provided to the researchers fully anonymized. Four pilot studies were conducted October 2018–May 2019 to produce reliable guidelines based on the RWF and developed via the raters’ comments and suggestions.

In Table 2, the first pilot study, four independent raters applied the initial RWF to the annotation of 20 sentences and then explained their ratings. From this, we recognized some ambiguity in the reflection indicators as formulated in the guidelines given to the raters. In the second pilot study, the three independent raters applied the modified RWF to 40 random sentences. The modified RWF enabled them to reach a consensus regarding the three levels and the seven indicators. Some minor

areas of the RWF guidelines were then refined. In the third and fourth pilot studies, two independent raters applied the RWF as framed after improvements. A kappa statistic (k) used to determine the inter-rater reliability and adjust for the possibility of a chance agreement between the coders. The inter-rater reliability of 0.87 and 0.78, respectively, which was substantial to almost perfect agreement (Landis & Koch, 1977).

Table 2: The inter-rater reliability computed for each iteration of the RWF during the four pilot tests

Date of the pilot test	#iteration	Sample	# raters	k
October 2018	1	20	4	0.52
January 2018	2	40	3	0.73
March 2019	3	100	2	0.87
May 2019	4	200	2	0.78

5 CONCLUSION AND FUTURE WORK

This research has answered two research questions that explored the characteristics of RW to identify the assessment indicators and the levels relating to RW in CS education. Based on the thematic analysis of the questionnaire, the RW framework was proposed; this has three levels and seven indicators to assess RW produced in the context of CS education. The future work will be using the findings to produce a labeled dataset to use it to develop an LA tool. That will automate RW analysis based on machine learning and rule-based approaches for determining the features of RW.

ACKNOWLEDGEMENT

The research was partially funded by Kuwait Foundation for the Advancement of Sciences (KFAS); Project code CB19-68SM-01; and Kind Saud bin Abdulaziz University for Health Sciences (KSAU-HS); scholarship program for Teacher Development, Department of Languages and Cultures.

REFERENCES

- Alrashidi, H., Joy, M., & Ullmann, T. (2019). *A Reflective Writing Framework for Computing Education*. Paper presented at the ITiCSE 2019: Proceedings of the 24th Annual ACM Conference on Innovation and Technology in Computer Science Education, Aberdeen, UK.
- Dorodchi, M., Benedict, A., Desai, D., Mahzoon, M. J., MacNeil, S., & Dehbozorgi, N. (2018). *Design and Implementation of an Activity-Based Introductory Computer Science Course (CS1) with Periodic Reflections Validated by Learning Analytics*. Paper presented at the 2018 IEEE Frontiers in Education Conference (FIE).
- Fekete, A., Kay, J., Kingston, J., & Wimalaratne, K. (2000). *Supporting reflection in introductory computer science*. Paper presented at the ACM SIGCSE Bulletin.
- George, S. E. (2002). Learning and the reflective journal in computer science. *Australian Computer Science Communications*, 24(1), 77-86.
- Gibson, A., Aitken, A., Sándor, Á., Buckingham Shum, S., Tsingos-Lucas, C., & Knight, S. (2017). *Reflective writing analytics for actionable feedback*. Paper presented at the Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17.
- Grossman, R. (2009). Structures for facilitating student reflection. *College Teaching*, 57(1), 15-22.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.

- Moskal, A. C. M., & Wass, R. (2019). Interpersonal process recall: a novel approach to illuminating students' software development processes. *Computer Science Education*, 29(1), 5-22. doi:10.1080/08993408.2018.1542190
- Stone, J. A., & Madigan, E. M. (2007). Integrating reflective writing in CS/IS. *ACM SIGCSE Bulletin*, 39(2), 42-45.
- Ullmann, T. D. (2019). Automated Analysis of Reflection in Writing: Validating Machine Learning Approaches. *International Journal of Artificial Intelligence in Education*, 1-41.
- Wong, F. K., Kember, D., Chung, L. Y., & Yan, L. (1995). Assessing the level of student reflection from reflective journals. *Journal of advanced nursing*, 22(1), 48-57.

Identifying negative language transfer in writing to increase English as a Second Language learners' metalinguistic awareness

Leticia Farias Wanderley

EdTeKLA Research Group, Department of Computing Science, University of Alberta
fariaswa@ualberta.ca

Carrie Demmans Epp

EdTeKLA Research Group, Department of Computing Science, University of Alberta
demmanse@ualberta.ca

ABSTRACT: This workshop paper describes the design and experimental application of an automated error detection tool that informs English as a Second Language (ESL) learners about possible negative language transfer effects in their writing. As these effects are known to hinder learners' proficiency in a second language, we hope to increase learners' metalinguistic awareness by explaining the latent causes of their misconceptions. The study utilizes an error annotated dataset of ESL learners' essays to validate common negative language transfer cases in different native languages and inform the error detection tool's feedback.

Keywords: writing analytics, language transfer, second language acquisition

1 INTRODUCTION AND BACKGROUND

English has become the *lingua franca* that is used in interactions between native speakers of distinct languages (Cenoz & Jessner, 2000). Most online content, business transactions, and international communications employ English. Given its spread and applicability, it is no wonder that people are interested in learning how to communicate in English. However, English learners struggle to acquire the language, especially when it differs from their mother tongues. In this work, we seek to provide feedback on learners' writing, elucidating errors that could be caused by discrepancies between English and their native languages. By doing so, we hope to raise learner awareness of the languages' grammatical distinctions and improve their writing skills.

When learning a novel language, learners often rely on their native languages' grammatical structures to form utterances in that new language. This phenomenon is known as language transfer, and it can be observed in all linguistic levels (Selinker, 1969). In the process of second language acquisition, language transfer is one of the strategies used by learners, intentionally and unintentionally, to communicate in that second language. This phenomenon usually occurs when the learners are unsure about the correct way to express themselves and can lead to them making grammatical errors due to a mismatch between rules in the two languages. The type and frequency of language transfer observed in learners' utterances vary according to their first language (L1) and proficiency in the second language (L2). The more proficient the learners, the more aware of L2 rules and their application they are. A less

proficient learner will rely on transfer more often. The nature of transfer observed, negative or positive, will depend on the amount of overlap between L1 and L2 rules. When grammatical rules between the L1 and L2 differ, the transferred language structure may result in an invalid utterance, according to the L2's rules. This negative language transfer effect can be defined as a lack of metalinguistic awareness in the L2 that forces learners to fall back on their L1s.

According to Gombert (1992), metalinguistic awareness is the ability to analyse language as an object, reflecting on its form and rules. Learners' metalinguistic awareness increases as they employ the language and receive feedback about their utterances. In this work, we set out to explore how error feedback, informed by the learners' L1s, can improve their L2 writing. We hope to increase the learners' L2 metalinguistic awareness by explaining the potential negative language transfer effects that lead to incorrect utterances.

2 LANGUAGE TRANSFER: AUTOMATING ITS DETECTION

The initial phase of this work consists of using automated methods to find negative language transfer evidence in English as a Second Language (ESL) learner writing. This step is essential to identify which aspects of each native language are transferred to English. Once negative language transfer evidence is established, we need to determine how to include it in the learner feedback and then measure whether it affects the learners' metalinguistic awareness and writing skills.

To automate negative language transfer detection, the writing assistant tool will leverage existing error correction systems, such as the LanguageTool¹ API, to identify and suggest corrections for errors in the learners' utterances. Our tool will then analyse the incorrect and corrected utterances' structures by identifying the grammatical categories, or parts-of-speech, of the words that form the utterances, in a process called part-of-speech tagging. Once the utterances' grammatical structures are found they can be compared against structures commonly used in the learner's L1.

The same part-of-speech tagging process will be applied to written corpora in the learner's native language to model a distribution of structures in that language. By comparing the frequency of the incorrect and corrected part-of-speech structures to part-of-speech sequences of the same length in the learner's native language, it is possible to determine whether the learners' incorrect utterances employ a commonly used sequence of parts-of-speech from their native languages and whether the error correction system's suggestion is valid in the learner's L1. If the learner's utterance structure, which is incorrect in English, is found to be common and valid in the learner's native language or the system corrected structure is not valid in the learner's L1, the system will flag a negative transfer error and provide feedback that references the relevant distinction between English and the L1. If there is not a clear correlation between the incorrect and corrected part-of-speech sequences in English and the distribution of sequences in the L1 part-of-speech tagged corpora (e.g., the incorrect utterance's structure used is not common in the learner's native language) the system will still provide error

¹ <https://languagetool.org/>

feedback. However, this feedback will not reference the learner's L1 nor will it reference language transfer effects.

To validate and test this methodology we will use a learner essay dataset from Cambridge's First Certificate in English exam (FCE). This error annotated dataset contains essays from 1,244 learners who, in total, have 16 different L1s (Yannakoudakis et al., 2011). The learners' errors are annotated and corrected following the error-tagging system described by Nicholls (2003). Each essay data point is not only annotated with writing errors and their respective correction but also complemented by metadata about the native language of its author and the score assigned by the annotator. The FCE dataset will be employed in the identification of common syntactic negative language transfer effects, as it contains the error annotated and corrected essays, and the learners' L1. This data will confirm the existence of language transfer and the tool's capacity to detect it. One of the advantages of using this dataset is the fact that it was manually tagged and corrected, hence there is no need to employ any extra error detection. It also contains essays that were designed to assess learners' writing skills in English at an intermediate level (Yannakoudakis et al., 2011), which enables the analysis of more complex syntactic structures. Another critical application of this dataset is in the language transfer detection model tuning, as the annotated errors can be applied to determine the distribution thresholds for what it means for a structure to be common in different languages.

Our negative transfer detection tool is under development. The current status of the development is that the FCE dataset has been processed and part-of-speech tagged. The errors have been grouped into syntactic and non-syntactic clusters. The next step of the implementation is to process corpora in different languages, initially Spanish and Chinese, to obtain the distribution of part-of-speech sequences necessary to identify language transfer effects.

3 PROPOSED STUDY: UNDERSTANDING THE SUPPORT TOOL'S EFFECT

To understand whether the negative language transfer feedback supports language learning and improves the learners' writing skills, we will recruit ESL learners into two groups. Both groups will use an online editor to write short essays in English. One group will receive negative language transfer informed feedback, while the other will receive generic feedback about the grammatical errors. Before and after the essay writing sessions, we will assess learner awareness of English syntax through cloze tests, in which they select the best option among negative language transfer and correct alternatives; error correction tasks, in which they correct ungrammatical sentences that contain negative transfer; and selection tasks, in which learners choose the grammatical version of a sentence from a set of options (Chireac et al., 2019).

In the proposed study the negative transfer feedback will be delivered by a web browser extension that is under implementation by a team of software developers. This extension will use the error correction API to detect errors in the learners' utterances as they write on an online editor. It will call attention to writing errors by highlighting the incorrect utterances in the users' texts. If the users wish to understand the nature of the error, they can select the highlighted text to make the extension display a pop-up box

containing an explanation about why the utterance is incorrect. For errors caused by negative language transfer, the feedback explanation will contrast specific rules in the user's L1 and English that are associated with the identified error. The extension users will be able to select their native languages to enable language transfer feedback.

As negative language transfer is a typical phenomenon in second language acquisition, we propose an automated error detection tool that will give learners more insight about writing errors that potentially arise from differences between their native languages and English. With that, we hope to support ESL learners' language development by giving them more opportunities to understand the causes of their errors. We expect that the addition of feedback based on language transfer effects will enhance learners' English metalinguistic awareness and, hence, their writing skills.

REFERENCES

- Cenoz, J., & Jessner, U. (Eds.). (2000). *English in Europe: The acquisition of a third language* (Vol. 19). Multilingual Matters.
- Chireac, S. M., Francis, N., & McClure, J. (2019). Awareness of Form and Pattern in Literacy Assessment: Classroom Applications for First and Second Language. *The Reading Matrix: An International Online Journal*, 19(1).
- Gombert, J. E., & Gombert, J. É. (1992). *Metalinguistic development*. University of Chicago Press.
- Nicholls, D. (2003, March). The Cambridge Learner Corpus: Error coding and analysis for lexicography and ELT. In *Proceedings of the Corpus Linguistics 2003 conference* (Vol. 16, pp. 572-581).
- Selinker, L. (1969). Language transfer. *General linguistics*, 9(2), 67.
- Yannakoudakis, H., Briscoe, T., & Medlock, B. (2011, June). A new dataset and method for automatically grading ESOL texts. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1* (pp. 180-189). Association for Computational Linguistics. <https://dl.acm.org/doi/10.5555/2002472.2002496>

Towards a Taxonomy of Writing Activities

Yoram M Kalman

The Open University of Israel
yoramka@openu.ac.il

Laura K Allen

University of New Hampshire
Laura.Allen@unh.edu

ABSTRACT: This workshop paper describes efforts to develop a taxonomy of writing activities that are involved in the production of a new text. The taxonomy under development is a first step towards the automated characterization of the typing and pointing activities which go into the writing and editing of texts on digital devices.

Keywords: Writing analytics; Taxonomy; Logger; Editing

1 WRITING AND EDITING ON DIGITAL DEVICES

The advent of digital devices and applications and the consequent proliferation of computer-based writing opens the opportunity for writing analytics scholars to study not only the product of the writing process (Shermis, Burstein, & Zechner, 2010), but also the process of writing (Van Waes, Leijten, Lindgren, & Wengelin, 2016). This endeavor has important implications for both theories of writing as well as real-world applications. A better understanding of how texts are produced and edited can improve our theoretical understanding of the cognitive processes involved in writing, the stages of text creation and development, individual differences amongst writers, and the linguistic properties associated with high-quality writing. These insights can in turn be applied to improve writing education, through the diagnosis and monitoring of specific writing features to the delivery of real-time feedback, and much more.

The primary tools used for studying writing processes on digital devices are loggers which track the keystrokes and pointing events that occur as individuals create texts. Basic loggers track only the keystrokes – i.e. the keys individuals hit as they create the text. More advanced loggers also record the evolving text in a detailed manner (*Draftback*, 2019; Kalman, Adam, & Blau, 2019). These loggers track every writing and editing activity, and their subsequent impact on the evolving text. Analyses of the output of these loggers have raised the need to create a taxonomy of the writing activities that individuals enact as they produce texts. Such a taxonomy will simplify the results of these loggers by interpreting the captured information and transforming it into specific named activities (e.g., typo correction). Consider cooking as an example -- it is simpler to state that "the onion was chopped" rather than describing how the cook took a knife, cut the onion top, split it into two halves, peeled the onion, lay the onion half on a cutting board, made many vertical slices etc.

In this workshop we describe our initial steps in creating a taxonomy of writing and editing activities. In particular, we will present the evolving taxonomy and discuss its usefulness, limitations, and our anticipated future directions.

2 A PRELIMINARY TAXONOMY OF WRITING ACTIVITIES

An ideal taxonomy of writing and editing activities would unambiguously classify all activities of a writer. Our work on developing this taxonomy relies on input from other taxonomies (e.g. in biology) and from classifications used in the teaching and analysis of writing (Van Waes, Leijten, Lindgren, & Wengelin, 2016). The development of the taxonomy has been iterative, relying on a trial and error method where classification criteria have been applied to essay writing samples (in English) in an effort to identify both redundancies and gaps. The work is still ongoing, and our most current version of the taxonomy will be presented in the workshop. Below we describe some of the primary features of the developing taxonomy.

The taxonomy characterizes activities that occur at three levels of the text: word, sentence and document. It classifies activities that occur at each level, while recognizing that as long as a writer is modifying a text at a lower level, it is futile to classify the activity at a higher level. For example, as long as I am still modifying a word (e.g. changing the word *home* to the word *house*), it is unnecessary to attempt to classify the modification I made to the sentence that contains that word. The taxonomy creates three "histories" which describe the full evolution of each unit (word, sentence, document) as it is created and as it evolves.

In the workshop we will demonstrate how the "histories" produced by the taxonomy help to characterize the evolution of each of the units. We will specifically present our work in progress on the taxonomy, let participants experiment with using the taxonomy, and discuss limitations and directions for future research.

REFERENCES

- Draftback*. (2019). Retrieved from <http://draftback.com/>
- Kalman, Y. M., Adam, E., & Blau, I. (2019). A Novel Writing Analytics Approach to Study Multiple Information Sources Integration by Students. *Companion Proceedings 9 Th International Conference on Learning Analytics & Knowledge (LAK19)*, 9–10. Retrieved from <http://wa.utscic.edu.au/wp-content/uploads/2019/03/Kalman-Adam-Blau-2019-A-Novel-Writing-Analytics-Approach-to-Study-Multiple-Information-Sources-Integration-by-Students.pdf>
- Shermis, M. D., Burstein, J., & Zechner, K. (2010). Automated essay scoring: Writing assessment and instruction. In *International encyclopedia of education* (pp. 20–26). Elsevier Science.
- Van Waes, L., Leijten, M., Lindgren, E., & Wengelin, Å. (2016). Keystroke logging in writing research: Analyzing online writing processes. In *Handbook of writing research, 2nd ed.* (pp. 410–426). New York, NY, US: The Guilford Press.

Writing Analytics to Support Integration of Multiple Texts

Authors: Jovita M. Vytasek, Alexandra Patzak, Philip H. Winne,
Simon Fraser University
Emails: jvytasek@sfu.ca, apatzak@sfu.ca, winne@sfu.ca

ABSTRACT: Presentation of a learning analytics model to support multiple text integration.

Keywords: Integration, Learning analytics, Multiple-source use

Writing affords learners opportunities to critically evaluate multiple sources of information and learn about current issues and problems. However, integrating information across diverse sources with potentially conflicting perspectives is complex as learners juggle reading, selecting useful material, writing notes, and drafting and revising a final product (Segev-Miller, 2007; Spivey & King, 1989). This iterative and recursive process plus variations in document properties makes it challenging to design instructional supports. Most provide strategy instruction without capturing how it is applied (Barzilai, Zohar & Mor-Hagani, 2018). Scaffolding learners as they apply strategies may improve integration. We propose an interactive writing analytic to facilitate multi-document synthesis using learners' pre-writing annotations and system guides for organizing ideas during the writing process.

1 DIGITAL TOOLS TO SUPPORT MULTIPLE DOCUMENT INTEGRATION

Digital support tools for text integration mainly offer visual representations (e.g. maps, tables) or instructor developed prompts (see Barzilai et al., 2018) where learners label links between concepts (Hilbert & Renkl, 2008), visually represent multiple sources (Kingsley et al., 2015), and compare texts (Cameron et al., 2017). Prompting learners to classify information, for example inviting learners to identify *claims*, *explanations* and *evidence* in multiple texts, improves integration (Barzilai & Ka'adan 2017). General guidance to integrate texts (e.g. Britt et al., 2004), metacognitive prompts (e.g. González-Lama et al., 2016) and content specific prompts (e.g. VanSledright, 2002) appear to support integration although main effects of such interventions can be masked when prompts are combined with other learning strategies. Building on this work, our analytic incorporates organization prompts with personalized feedback to offer more directed, content specific guidance.

2 MODEL DRIVEN ANALYTIC DESIGN

Integrating content requires transforming information from multiple sources into a coherent product (Segev-Miller, 2007; Spivey & King, 1989). This involves (1) selecting information, (2) organizing selections to form a mental model, and (3) connecting/transforming selections to synthesize a coherent essay. The proposed analytic aids learners in all three tasks. Building on nStudy's advanced annotation system (Winne et al., 2019), we scaffold learners with a three-featured tool. First, learners are guided to tag and annotate selections in texts, and rate source quality. Tags are *main idea* and *support* plus a field to add keywords. Tagged selections can be linked as clusters of content. A mirroring analytic per source reflects number of selections by tag and a

keyword summary. Second, to refine main ideas, learners work in a graphical tool to group selections. Using text similarity metrics, the system offers an initial organization, clustering all selections, colour-coded by source. Learners then rearrange this draft map, add and remove selections, and create new clusters (see figure 1). Links between selections and clusters can be labeled to aid integration and linguistic transformations (Hilbert & Renkl, 2008). Map analytics describe the proportion of selections from high-quality sources and density of links to main ideas. Automated prompts highlight main ideas with few connections and rarely used high-quality sources. Third, learners drag selections and clusters from the map into a writing editor. Used map selections and clusters are marked. All three features work together simultaneously to promote productive reading, re-reading, drafting, and revision.

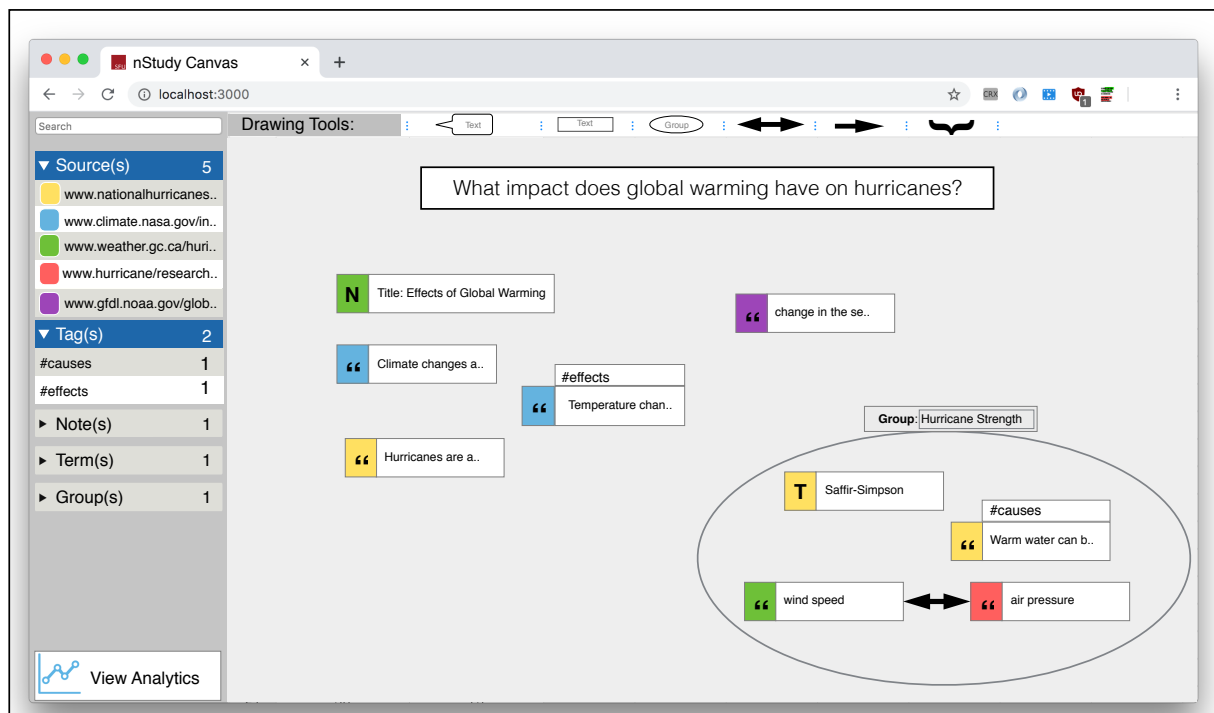


Figure 1 Student customized canvas map of annotations made to source materials

Future directions for research include investigating how to provide more tailored recommendations to learners to help them organize and locate source material they previously annotated, and related material they have yet to consult. As a first step, we suggest exploring three techniques for generating this feedback. First, learners operationalize a schema for classifying information by tags they assign. Adding keywords to tagged content and grouping terms, text selections, and notes in the canvas adds a second dimension for classifying information by disciplinary content. These meta-data about information learners process provide a basis for recommendations. When learners use an artifact in their draft essay, the system can remind learners of related information artifacts based on tags and groupings. Learners with low prior knowledge, or who suffer high cognitive load or a production deficiency (overlooking opportunity to use information due to absent or ill-formed standards for metacognitive monitoring) can be guided to consider and organize material distributed across multiple texts (Bråten, Anmarkrud, Brandmo, & Strømsø, 2014). Learners also are afforded opportunity to extend search throughout the lattice of information artifacts by examining tags and groupings of system-recommended items. A second method for supporting learners' consideration

of the conceptual structure of a corpus and a draft essay is a network visualization of terms. The network can be constructed on the basis of co-occurrences of terms such as when a term is defined using other terms and co-occurrences of terms within easily identified grammatical units (e.g., sentences, paragraphs). Co-occurrences express conceptual relationships among key concepts. A learner can survey a termnet summarizing key content in a source and how the source relates those concepts as a comparison to the graph representing the conceptual structure of their draft. For example, if term A is linked to term C through term B, but the term B is missing in the draft, adding this term may be an important revision to establish logical structure or coherence. Graphs also could be leveraged as a method to search for sources that are “geometrically” similar, to reinforce and coherently elaborate content in a draft, and sources that are distinct, to test for bias in a draft. Visually decorating learner-created groups and tags could add dimensionality representing their conceptual structuring of material and suggesting potential re-conceptualization. Research shows graphic organizers aid assembling information from multiple texts, guide judicious review of intertextual connections, and prompt categorization and linking processes (Barzilai, et al., 2018). Research is needed to explore termnets as digital graphic organizers to support these processes. A third line of research might explore helping learners generate queries of sources. Supplying learners with tags such as *hypothesis*, *method*, *result*, *exception*, etc. operationalizes a model of rhetorical classes overlapping with requirements for essays. Intersecting classes and terms, such as *hypothesis* × “wind speed,” is a tool learners could use to more efficiently locate and compare content in sources and their evolving draft. This may benefit learners to identify and select content as they construct a representation of how concepts are used in and across texts (Britt & Sommer, 2004).

Amalgamating these features affords learners opportunities to experiment with and tune multiple learning strategies to achieve their goals. As learners tag and annotate sources, map artifacts, and search, trace data they generate reflects cognitive and metacognitive processes they use to organize particular content (Winne & Marzouk, 2019). Traces generated as learners used these tools open windows onto their self-regulated learning (SRL) (Winne, 2018) and, as big data are amassed, for bootstrapping each learner’s skills and broadening foundations for learning science (Winne, 2017).

Novel research opportunities arise by integrating the proposed system with other writing analytics to provide feedback about learners’ pre-writing processes and revision activities. Refining data that describes conditions precipitating revisions and actions taken to realize revisions will lead to insights about timing and personalizing prompts. Through cycles of engagement on multiple assignments, it is possible to observe how strategies evolve as learners receive and act on prompts, and gradually internalize metacognitive skills that improve skills in integrating concepts from multiple texts (Barzilai & Ka’adan, 2017).

We invite comments on our model and seek research collaborations. Capturing trace data on learners’ engagements with and organization of source information for essays offers unique opportunities to research learning processes, self-regulation and the utility of intervention prompts.

REFERENCES

Bråten, I., Anmarkrud, Ø., Brandmo, C., & Strømsø, H. I. (2014). Developing and testing a model of direct and indirect relationships between individual differences, processing, and multiple-text comprehension. *Learning and Instruction*, 30, 9–24.

Creative Commons License, Attribution - NonCommercial-NoDerivs 3.0 Unported (CC BY-NC-ND 3.0)

- Barzilai, S., & Ka'adan, I. (2017). Learning to integrate divergent information sources: the interplay of epistemic cognition and epistemic metacognition. *Metacognition and Learning*, 12(2), 193–232.
- Barzilai, S., Zohar, A. R., & Mor-Hagani, S. (2018). Promoting integration of multiple texts: a review of instructional approaches and practices. *Educational Psychology Review*, 30(3), 973-999.
- Britt, M. A., & Rouet, J. F. (2012). Learning with multiple documents: component skills and their acquisition. In J. R. Kirby & M. J. Lawson (Eds.), *Enhancing the quality of learning: dispositions, instruction, and learning processes*. New York: Cambridge University Press.
- Britt, M. A., & Sommer, J. (2004). Facilitating textual integration with macro-structure focusing tasks. *Reading Psychology*, 25(4), 313–339.
- Cameron, C., Van Meter, P., & Long, V. (2017). The effects of instruction on students' generation of self-questions when reading multiple documents. *Journal of Experimental Education*, 85(2), 334–351.
- González-Lamas, J., Cuevas, I., & Mateos, M. (2016). Arguing from sources: design and evaluation of a programme to improve written argumentation and its impact according to students' writing beliefs. *Children and Learner*, 39(1), 49–83.
- Hilbert, T. S., & Renkl, A. (2008). Concept mapping as a follow-up strategy to learning from texts: what characterizes good and poor mappers? *Instructional Science*, 36(1), 53–73.
- Segev-Miller, R. (2007). Cognitive processes in discourse synthesis: the case of intertextual processing strategies. In M. Torrance, L. Van Waes, & D. Galbraith (Eds.), *Writing and cognition* (pp. 231-250). Bingley: Emerald Group.
- Spivey, N. N., & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly*, 24(1), 7–26.
- VanSledright, B. A. (2002). Confronting history's interpretive paradox while teaching fifth graders to investigate the past. *American Educational Research Journal*, 39(4), 1089–1115.
- Winne, P. H. (2017). Leveraging big data to help each learner upgrade learning and accelerate learning science. *Teachers College Record*, 119(3), 1-24.
- Winne, P. H. (2018). Cognition and metacognition within self-regulated learning. In D. Schunk & J. Greene (Eds.), *Handbook of self-regulation of learning and performance*. (2nd ed., pp. 36-48). New York, NY: Routledge.
- Winne, P. H., & Marzouk, Z. (2019). Learning strategies and self-regulated learning. In J. Dunlosky & K. Rawson (Eds.), *Cambridge handbook of cognition and education* (pp. 696-715). New York, NY: Cambridge University Press.
- Winne, P. H., Teng, K., Chang, D., Lin, M. P.-C., Marzouk, Z., Nesbit, J. C., Patzak, A., Raković, M., Samadi, D., & Vytasek, J. (2019). nStudy: Software for Learning Analytics about Processes for Self-Regulated Learning. *Journal of Learning Analytics*, 6(2), 95–106.

Argument component identification and its application in feedback on Dutch essays

Liqin Zhang, Howard Spoelstra
Open University of the Netherlands
liqin.zhang,howard.spoelstra@ou.nl

Marco Kalz
Open University of the Netherlands, Heidelberg University of Education
kalz@ph-heidelberg.de

ABSTRACT: Assessment for feedback on argumentative essays is challenging. However, current research on argument mining opens possibilities to developing automated argumentation assessment tools. Currently these tools are mainly available for the English language. Additional effort is required to adapt the techniques to e.g. Dutch. This study focuses on argument component identification on Dutch essays and how to present the results so they can be used as formative feedback.

Keywords: argument mining; Dutch essay; formative feedback

1 INTRODUCTION

Writing argumentative essays is challenging to both students and teachers. A survey among teachers (Authors, submitted) shows that both difficulty and time consumption for providing feedback on arguments are high. While automated assessment of argumentation and formative feedback generation would be able to alleviate the pressure on teachers, it is still not commonly available. On the one hand, argument mining, whose goal is to extract argumentation features, is still emerging (Lippi & Torroni, 2016), on the other hand, the state-of-the-art technology is mostly developed in the context of English, meaning that the application of the methods to other languages, such as Dutch, is restricted. We aim to develop a model to support argumentation analysis of essays in Dutch that can provide formative feedback as both quality indication and guidelines for improvement. Specifically, an argumentation component identification model for Dutch argumentative essays is developed by adopting previous work based on the English language (Stab & Gurevych, 2017a). The model is expected to function as an argumentation analysis component in a Dutch writing analysis tool. Based on the affordances of the model we explore the possible approaches to generate formative feedback.

2 RELATED STUDIES

2.1 Argument mining and formative feedback

Previous studies have explored the application of argument mining for essay assessment in terms of holistic essay quality and argumentation quality (Ghosh, Khanam, Han, & Muresan, 2016; Wachsmuth et al., 2017; Toledo et al., 2019). Additionally, giving effective formative feedback is already one of the main challenges in the current development of automated essay assessment (Strobl et al., 2019),

which is particularly true for formative feedback on the argument components in texts. The study of applying argument mining for relevant feedback generation is still rare. One of the few studies into this matter is by Stab & Gurevych (2017b), who attempt to provide feedback on arguments according to the output of an argumentation analysis model (Stab & Gurevych, 2017a). Only recent studies into argument mining began to focus on developing such models aiming to translate the results of analysis into useful formative feedback for further improvement of argumentative writing. For instance, Carlile et al. (2018) and Gao et al. (2019) developed rubric-based corpora and taxonomies which are used to classify various argumentation features, such as persuasiveness, coverage, coherence etc. into quality levels and it is beneficial for further studies of providing feedback based on the classification.

2.2 Argument mining in multilingual settings

Because of the lack of the argument mining research in non-English context, there are few human-annotated argumentation corpora available for multilingual argument mining research. Since the workload to create such human-annotated corpora is heavy, Eger et al. (2018) explore an alternative approach which involves machine translation and tag projection in order to extend existing models to afford multilingual application. The approach is as follows: A human-annotated L1 corpus (source language) with tags for major claims, claims, and premises is translated to L2 (target language) using Google Translate. By using fast-align (Dyer, Chahuneau, & Smith, 2013), each token in L1 is then aligned to the corresponding parallel text in L2. As each token in L1 is annotated by the aforementioned tags, these tags from L1 are projected¹ to the corresponding aligned tokens in L2. Eger et al. (2018) have shown that using such a machine-generated corpus to train a model for argument component identification performs comparably to the models trained by using a human-annotated corpus. This approach is also successfully applied to identify argument component relations for other languages, such as Portuguese (Rocha, Stab, Cardoso, & Gurevych, 2019).

3 ARGUMENT COMPONENT IDENTIFICATION FOR DUTCH

The identification of argument component identification is one of the basic tasks in argument mining. This study starts with developing an argumentation component identification model for Dutch, based on the studies of Stab & Gurevych (2017a) and Eger et al. (2018).

3.1 Data

Human-annotated argumentation corpora for Dutch argument mining are not available. To create a corpus of Dutch essays with argumentation structure annotations, a human-annotated corpus of 402 essays in English (Stab & Gurevych, 2014) is translated into Dutch with Google Translate. Afterwards, following the approach by Eger et al. (2018) the tags from the essays in English are projected into the Dutch texts. As a result, we obtain a corpus in Dutch containing 402 persuasive essays, annotated with argumentation components (major claim, claim, and premise) on the token-level. The annotations are in Inside-outside-beginning (IOB) format (Ramshaw & Marcus, 1999).

¹ For the detail of the projection method please see Eger et al. (2018).

3.2 Implementation of the model

When a training dataset is available in which the labels are on the token level, the implementation of a model is a task of token-level sequence tagging. The model is a bidirectional Long-Short-Term-Memory (LSTM) neural network with a CRF layer. The parameters set to train the model are derived from Eger et al. (2018). The model is trained in 5 runs, with 50 epochs in each run. The Dutch words are represented with 300-dimension vectors which is trained on the Wikipedia Dutch articles using a skip-gram model (Bojanowski et al., 2017). The framework of Kahse (2018) is applied for the implementation of the model. The adapted Dutch data is split into train/develop/test sets, and the performance of the model is evaluated by precision/recall/F1 scores.

4 PRESENTATION OF FORMATIVE FEEDBACK

Based on the affordances of the model, we propose to provide formative feedback which we expect to be useful for teachers and students for them to improve their performance in writing argumentatively. As the current model identifies the argument components and types, it is feasible to highlight them in the text. Descriptive statistics (such as the number of premises and claims) can also be provided as feedback, as these relate to argument quality. Visualizing the argumentation structure (see Figure 1) is also effective to provide insight into the argumentation structure of the essay (Chiang, Fan, Liu, & Chen, 2016).

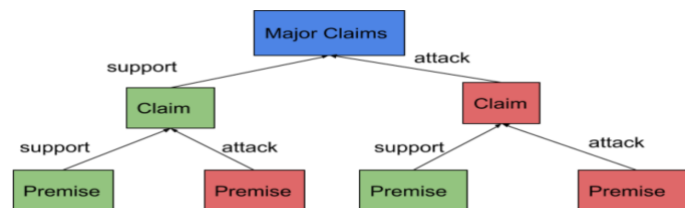


Figure 1: An example of visualization of argumentation structure.

5. LIMITATIONS AND FUTURE WORK

The possibility of generating effective formative feedback depends on the performance of the argumentation component identification model. The performance of the model needs to have a sufficiently accurate to use it for formative feedback purposes. Meanwhile, the variety in the formative feedback is possibly limited by the affordances of the model. In the future word we will therefore focus on the assessment of the reliability of the approach and the different ways to present feedback based on results of the analysis. We suggest that the model can be improved by analyzing the relations between argument components with assessment rubrics, allowing the student to understand their performance. Last but not the least, the usefulness of the formative feedback generated by the argument analysis approaches should be evaluated by conducting an empirical experiment involving students and teachers.

REFERENCES

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5, 135-146.

- Carlile, W., Gurrapadi, N., Ke, Z., & Ng, V. (2018, July). Give me more feedback: Annotating argument persuasiveness and related attributes in student essays. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 621-631).
- Chiang, K. H., Fan, C. Y., Liu, H. H., & Chen, G. D. (2016). Effects of a computer-assisted argument map learning strategy on sixth-grade students' argumentative essay reading comprehension. *Multimedia Tools and Applications*, 75(16), 9973-9990.
- Dyer, C., Chahuneau, V., & A Smith, N. (2013). A simple, fast, and effective reparameterization of ibm model 2.
- Eger, S., Daxenberger, J., Stab, C., & Gurevych, I. (2018). Cross-lingual Argumentation Mining: Machine Translation (and a bit of Projection) is All You Need!. *arXiv preprint arXiv:1807.08998*.
- Gao, Y., Driban, A., McManus, B. X., Musi, E., Davies, P., Muresan, S., & Passonneau, R. J. (2019, August). Rubric Reliability and Annotation of Content and Argument in Source-Based Argument Essays. In *Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications* (pp. 507-518).
- Ghosh, D., Khanam, A., Han, Y., & Muresan, S. (2016, August). Coarse-grained argumentation features for scoring persuasive essays. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* (pp. 549-554).
- Kahse, T. (2019). Multi-task learning for argumentation mining. *arXiv preprint arXiv:1904.10162*.
- Lippi, M., & Torroni, P. (2016). Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology (TOIT)*, 16(2), 10.
- Ramshaw, L. A., & Marcus, M. P. (1999). Text chunking using transformation-based learning. In *Natural language processing using very large corpora* (pp. 157-176). Springer, Dordrecht.
- Rocha, G., Stab, C., Cardoso, H. L., & Gurevych, I. (2018, November). Cross-Lingual Argumentative Relation Identification: from English to Portuguese. In *Proceedings of the 5th Workshop on Argument Mining* (pp. 144-154).
- Stab, C., & Gurevych, I. (2014, August). Annotating argument components and relations in persuasive essays. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers* (pp. 1501-1510).
- Stab, C., & Gurevych, I. (2017a). Parsing argumentation structures in persuasive essays. *Computational Linguistics*, 43(3), 619-659.
- Stab, C., & Gurevych, I. (2017b). Training Argumentation Skills with Argumentative Writing Support. *Computational Linguistics*, 43(3), 619-659.
- Strobl, C., Ailhaud, E., Benetos, K., Devitt, A., Kruse, O., Proske, A., & Rapp, C. (2019). Digital support for academic writing: A review of technologies and pedagogies. *Computers & Education*, 131, 33-48.
- Toledo, A., Gretz, S., Cohen-Karlik, E., Friedman, R., Venezian, E., Lahav, D., ... & Slonim, N. (2019). Automatic Argument Quality Assessment--New Datasets and Methods. *arXiv preprint arXiv:1909.01007*.
- Wachsmuth, H., Naderi, N., Hou, Y., Bilu, Y., Prabhakaran, V., Thijm, T. A., ... & Stein, B. (2017, April). Computational argumentation quality assessment in natural language. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers* (pp. 176-187).

Third Workshop on Social and Emotional Learning (SEL): Integrate SEL and Learning Analytics

Yuan Wang

EdPlus Action Lab, Arizona State University,
elle.wang@asu.edu

Maria Ofelia San Pedro, Jason Way, John Whitmer
ACT Inc.

{sweet.san.pedro, jason.way, john.whitmer}@act.org

Srecko Joksimovic

University of South Australia
srecko.joksimovic@unisa.edu.au

ABSTRACT: Following the success of hosting the SEL workshops at LAK18 and LAK19, we propose to organize the third SEL workshop at LAK20 by focusing on connecting assessment and analyses on SEL attributes with learning analytics (LA) methods that tap into the increasingly rich learner behavioral data. Toward expanding this focal area and echoing the theme of LAK20 on “shaping the future of the field”, we invite researchers and practitioners to systematically review and share learning analytics-driven methods into evaluating SEL attributes toward improving student learning and beyond. In this proposal, we introduce the significance and history of the SEL workshop series, followed by objectives and proposed organizational plans.

Keywords: Learning Analytics, social and emotional learning, non-cognitive assessments

1 WORKSHOP BACKGROUND

1.1 Motivation & Significance

The importance of fostering and measuring non-cognitive or social and emotional learning (SEL) skills, commonly viewed as critical personal attributes necessary for success in classrooms, the labor market, and life in general, has been widely recognized (Duckworth & Yeager, 2015; Heckman, Stixrud & Urzua, 2006) with an increasing need of assessing them at scale (Buckingham Shum & Crick, 2016). Learning analytics (LA) as one of the fastest-growing emergent disciplines have shown great potential in applying large-scale learner behavioral data, such as those derived from clickstream, toward the goal of optimizing learning (Siemens & Long, 2011; Ferguson, 2012). Shum and Ferguson proposed “social learning analytics” as a subset of the LA while acknowledging the importance of participatory culture in online learning (2012). Within this focal area, many researchers have applied social network analysis into understanding how network formation among learners influence learning (e.g., Gašević et al, 2019). We would like to expand

upon this concept and aim to expand on the area of social learning analytics by incorporating LA-driven analyses into individual-level social and emotional constructs such as those related to personality and learning habits.

1.2 SEL (Definition, recent development, etc.)

There are a great variety of terms that people use when referring to skills that fall outside of the traditional math and verbal cognitive skills: 21st Century skills, life skills, soft skills, noncognitive skills, social and emotional learning skills, personal skills, character, etc. For this workshop, we adopt the term social and emotional learning (SEL) and the following definition from John & DeFruyt, 2015: “individual capacities that (a) are manifested in consistent patterns of thoughts, feelings, and behaviours, (b) can be developed through formal and informal experiences, and (c) influence important socioeconomic outcomes throughout an individual’s life” (p.4).

SEL skills are both related to important outcomes across the lifespan and develop over time. Outcomes associated with SEL skills include, but are not limited to, academic performance (Poropat, 2009), college retention (Robbins et al., 2004), behavioral problems (Ge & Conger, 1999), etc. Roberts, Walton, and Viechtbauer (2006) provide meta-analytic data demonstrating that SE skill development occurs naturally throughout the entire lifespan. Meta-analytic data also suggest that interventions can lead to significant SE skill change (Roberts et al., 2017) and suggest that school-based social and emotional learning programs can be effective for improving SE skills (Corcoran, Cheung, Kim, & Xie, 2018; Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Taylor, Oberle, Durlak, & Weissberg, 2017; Wiglesworth et al., 2016).

1.3 Development of the workshop (connecting back with previous years)

In the inaugural workshop held at LAK18 (Wang et al., 2018), the workshop focused on SEL assessment at scale while aiming at bridging the scarcity of scalable SEL skills assessments in digital environments; the second workshop at LAK19 (Wang et al., 2019) further enriched the goal by incorporating the goal of promoting inclusion and diversity. As an immediate follow-up to the second workshop, the workshop chairs proposed and received approval to edit a *Journal of Learning Analytics* special issue focusing on the assessment of 21st century skills. The third workshop strives to systematically examine how learning analytics methods can be used to evaluate SEL, and conversely, how understanding of student SEL skills can improve learning analytics models and the impact of interventions driven by these models.

2 RELEVANCE TO LAK & OBJECTIVES

In line with LAK2020’s theme of “shaping the future of the field”, we propose to expand on this area of social learning analytics by 1) offering tutorials to highlight existing work that utilizes various learning analytics methods into measuring and predicting critical SEL attributes; & 2) connecting researchers and practitioners to collaboratively advance and expand this field by envisioning future activities and development as a deliverable of this workshop. Toward these general goals, we propose the following objectives for this workshop:

3 OBJECTIVES

- 3.1 Evaluate existing studies that has been conducted in SEL that is relevant to learning analytics research and the directions suggested by that work (as well as gaps that can be addressed).
- 3.2 Identify promising methodological approaches to using learning analytics to assess SEL and estimate the level of effort and difficulty of using these approaches.
- 3.3 Create a “short list” of suggested approaches and research opportunities for participants to collaborate around and promote to the broader LAK community.

4 ORGANIZATIONAL PLANS

Type of event: Interactive workshop session; proposed schedule and duration: Half-day.

Type of participation: Mixed participation. The proposed workshop plans to elucidate a clearer picture of what SEL is, the different ways SEL has been talked about, and how learning analytics can facilitate in their assessments. Invited experts or researchers who are involved in projects related to the workshop’s theme of learning analytics methods to evaluate SEL constructs will present their work. And workshop participants will be encouraged to participate in interactive discussion (both verbal, group work and/or collaborative document writing) of how the workshop theme applies to their current work or interests.

Proposed activities: Presentations by organizers who are experts in LA; Invited presentations of work as related to the workshop them; guided small-group discussions, collaborative writing and presentations by the workshop participants.

Expected participant numbers: 25-30

Website: We also plan to integrate all relevant resources and contact information, produced before and during the workshop such as presentation slides and discussion notes, on the website to encourage ongoing communication and collaboration after the workshop.

Social Media: We will use #SELatLAK20 as the primary hashtag to encourage discussions and communication through Twitter.

Required equipment for the workshop: A conference room with a capacity for up to 40 people with a setup that allows for small group discussions. A computer and a screen for presentations are also needed.

REFERENCES

- Buckingham Shum, S., & Crick, R. D. (2016). Learning Analytics for 21st Century Competencies. *Journal of Learning Analytics* , 3 (2), 6-21.
- Corcoran, R. P., Cheung, A., Kim, E., & Xie, C. (2018). Effective universal school-based social and emotional learning programs for improving academic achievement: A systematic review and meta-analysis of 50 years of research. *Educational Research Review*, 25, 56-72.
- Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher* , 44 (4), 237-251.

- Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random House Incorporated
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82, 405-432.
- Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6), 304-317.
- Gašević, D., Joksimović, S., Eagan, B. R., & Shaffer, D. W. (2019). SENS: Network analytics to combine social and cognitive perspectives of collaborative learning. *Computers in Human Behavior*, 92, 562-577.
- Ge, X., & Conger, R. D. (1999). Adjustment problems and emerging personality characteristics from early to late adolescents. *American Journal of Community Psychology*, 27, 429-459.
- Heckman, J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24(3), 411-482.
- John, O., & De Fruyt, F. (2015). *Framework for the Longitudinal Study of Social and Emotional Skills in Cities*. Paris, France: OECD Publishing.
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, 135, 322-338.
- Robbins, S. B., Lauver, K., Le, H., Davis, D., Langley, R., & Carlstrom, A. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130, 261-288.
- Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychological Bulletin*, 143, 117-141.
- Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychological Bulletin*, 132, 1-25.
- Shum, S. B., & Ferguson, R. (2012). Social learning analytics. *Journal of educational technology & society*, 15(3), 3-26.
- Taylor, R., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development*, 88, 1156-1171.
- Wiglesworth, M., Lendrum, A., Oldfield, J., Scott, A., ten Bokkel, I., Tate, K., & Emery, C. (2016). The impact of trial stage, developer involvement and international transferability on universal social and emotional learning programme outcomes: A meta-analysis. *Cambridge Journal of Education*, 46, 347-376.
- Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, 46(5), 30.
- Wang, Y., San Pedro, M.O.Z., Joksimovic, S., & Way, J. (2019). Workshop on social emotional learning, assessment toward diversity and inclusion. *Proceedings of the 9th International Conference on Learning Analytics and Knowledge*.
- Wang, Y., San Pedro, M.O.Z., Joksimovic, S., Mckinnis, T., & Way, J. (2018). Workshop on non-cognitive assessments at scale for online learning. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*.

Addressing Drop-Out Rates in Higher Education

François Bouchet, Vanda Luengo

Sorbonne Université, France
{francois.bouchet, vanda.luengo}@lip6.fr

Geoffray Bonnin, Anne Boyer, Armelle Brun

Université de Lorraine, France
{geoffray.bonnin, anne.boyer, armelle.brun}@loria.fr

Mohamed Amine Chatti

University of Duisburg-Essen, Germany
mohamed.chatti@uni-due.de

Irene-Angelica Chounta

University of Tartu, Estonia
chounta@ut.ee

María Jesús Rodríguez-Triana, Kairit Tammet

Tallin University, Estonia
{mjrt, kairit}@tlu.ee

Agathe Merceron, Petra Sauer

Beuth University of Applied Sciences
{merceron, sauer}@beuth-hochschule.de

ABSTRACT: This proposal describes the goal and activities of the LAK 2020 half-day symposium on Addressing Drop-Out Rates in Higher Education (ADORE 2020). The purpose of the symposium is to bring together a community of stakeholders (namely, researchers and practitioners) who work on data-driven, learning analytics for detecting students at-risk and on strategic designs for addressing dropouts in Higher Education. Our goal is to promote knowledge sharing by building a knowledge base of successful practices and to communicate lessons learnt from the design and adoption of institutional analytics in diverse contexts in order to contribute to robust, sustainable and transferable analytical solutions.

Keywords: dropouts, higher education, institutional analytics, data-driven decision making.

1 SYMPOSIUM BACKGROUND

The working environment is constantly evolving. The labor market desperately needs graduates from different disciplines and also requires workers to keep themselves up to date, engaging in lifelong learning solutions (UNESCO, 2016). In such a scenario, Higher Education (HE) institutions play a crucial role. As several international reports show (European Commission, 2015; European Commission, 2017; EDUCAUSE, 2019), the educational community and its policy makers are concerned with the HE success rates, and try to find strategies to attract students to education, keep them on board, and guide them to successfully acquire their degrees.

Student dropout is a complex topic, which is affected by different personal, instructional, social and organisational factors. We focus mainly on instructional factors such as gaps in course and program designs, students' under-performance, absence of feedback loops and we seek possibilities to address these. The organizers of this symposium are exploring data-driven strategies to promote student retention, to provide post-entry support, guidance and counselling of students, and to scaffold students' meta-cognitive strategies. However, despite the potential of the ongoing research in supporting student's academic success, our analytical solutions are still in an early stage or piloting phase and, only a small number of stakeholders (mainly researchers) have access to them.

This symposium will focus on learning analytics approaches for reducing student dropout in HE. Our aim is to contribute to *"shaping the future of the field"* (LAK 2020 conference theme¹) by bringing together established research practices from various contexts (that is, different countries, different academic institutions and different domains), building a knowledge base of successful paradigms (for example, analytical approaches and decision-making strategies) and sharing the lessons learnt during the process of addressing student dropouts in Higher Education. The goal of this symposium is twofold:

- a) To create a community of stakeholders in order to share expertise, receive feedback and communicate lessons learnt from the design, adoption and application of data-driven practices (institutional analytics) for addressing dropouts in Higher Education, and;
- b) To contribute in building a knowledge base of successful practices that are essential for the adoption of learning and institutional analytics.

This symposium will emphasize the design and adaptation of robust, sustainable and transferable strategies for the future. To that end, we aim to report and guide each other in the following directions:

- Defining a solid basis for ethics, data privacy and compliance for the European General Data Protection Regulation (GDPR);
- Integrating the stakeholders in the loop and putting the students in the center;
- Promoting a holistic approach where reducing dropout is not only an institutional matter but a shared goal among stakeholders;
- Closing the loop to assess and provide evidence about the added value that strategies have in terms of user acceptance and impact on reducing dropout rates.

For any of the above themes, we welcome the contributions of researchers and practitioners. Contributions can take the form of papers for presentation (maximum 6 pages), posters or demos (maximum 3 pages).

2 ORGANISATIONAL DETAILS

2.1 Type of Event

Mini-tracks/Symposia. We aim at a program committee of about 20 members so that the review load should be one / two contributions maximum per reviewer.

¹ <https://lak20.solaresearch.org/>

2.2 Proposed Schedule and Duration

This symposium is planned as a half-day event. We propose the following schedule:

- 8:30am-8:40am: Welcome, introduction, and goal of the workshop.
- 8:45am-9:00am: Attendees present themselves shortly.
- 9:00am-10:30am: 4 x presentation + discussion (10 minutes presentation, 10 minutes discussion each).
- 10:30am-10:45am: Break (15 minutes).
- 10:45am-11:15am: Poster/Demo session (max. 6) with discussion or with handouts.
- 11:15am-12:00am: Discussion: shaping best practices and building a knowledge base
- 12:00am-12:30 am: Wrap up & dissemination of results & future joint actions & Goodbye.

2.3 Type of Participation

The event supports mixed participation. Both participants with a paper submission and interested delegates may register to attend.

2.4 Symposium Activities

The symposium will host paper presentations along with poster and demo sessions. Additionally, we aim to engage participants in semi-structured, round table discussions regarding ways to address dropouts in HE and specifically on the following directions: 1) student-centered, participatory design, 2) generalizability and transferability, 3) ethics and data privacy and 4) impact and added-value.

2.5 Expected participant numbers and planned dissemination activities to recruit attendants

The symposium aims at 20 participants. To recruit participants, we will communicate this event using social media platforms (Twitter, ResearchGate etc.) and mailing lists of international (SOLAR, EDM, EATEL, ISLS) and national (e.g., nordicLASI, SNOLA in Spain, ATIEF in France, GI in Germany) communities and initiatives. Additionally, we will launch a workshop website that will be linked to the LAK2020 website and we will form a program committee of about 20 members to disseminate the workshop further with the networks of the members.

It should be noticed that the organisers come from six different academic institutions from three countries, and at least the attendance of representatives from these institutions is guaranteed.

2.6 Required equipment

Projector, flipcharts, post-it notes, apple adapter and a room suitable for group discussion. The seating should first be arranged as rows.

3 ORGANISATIONAL DETAILS

The workshop goals are:

- Report and share among the participant experience with institutional LA solutions. How LA is supporting learning and success?

- Familiarize participants with different existing institutional LA solutions to address drop-out;
- Identify challenges and good practices;
- Bring together researchers, practitioners, educational developers and policymakers.

This will allow:

- For novice participants, to learn about the field and get involved;
- For more expert participants, to share their experiences and receive feedback;
- To facilitate interdisciplinary collaboration among the participants from different backgrounds like governance, researchers, teachers, and so on.

In this way, we aim to advance the field and discuss challenges and issues related to the institutional LA and student dropout. All accepted contributions will be published in the “LAK Companion Proceedings”. The outcomes of the workshop will be published on the workshop’s website. A further intended outcome is the joint publication of a handbook (with extended contributions from the participants) that will report and reflect on the symposium’s contributions and discussions as well as on envisioning the future of institutional analytics.

4 STRUCTURE AND CONTENT OF THE SYMPOSIUM WEBSITE

- Call for papers (theme, submission guidelines)
- Important Dates
- Workshop description
- Organizers, program committee
- Accepted papers
- Outcomes from the workshop after the workshop

REFERENCES

- UNESCO. (2016). *Education 2030: Incheon Declaration and Framework for Action for the implementation of Sustainable Development Goal 4: Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all*. Retrieved from: <https://goo.gl/xM5VW7>
- European Commission. (2015). *Dropout and completion in Higher Education in Europe*. Retrieved from: <https://goo.gl/CedMcF>
- European Commission. (2017). *Renewed EU agenda for higher education*. Retrieved from: <https://goo.gl/yt14r6>
- EDUCAUSE. (2019). *EDUCAUSE Horizon Report 2019 Higher Education Edition*. Retrieved from: <https://library.educase.edu/-/media/files/library/2019/4/2019horizonreport.pdf>

Accuracy of a Cross-Program Model for Dropout Prediction in Higher Education

Kerstin Wagner, Agathe Merceron, Petra Sauer

Beuth University of Applied Sciences Berlin

{kerstin.wagner, sauer, merceron}@beuth-hochschule.de

ABSTRACT: Reducing dropout rates in higher education would allow increasing the number of graduates. If one can predict early enough whether a student might drop out, targeted counseling could be put in place. This work replicates the approach of Berens et al. (2019) to predict whether students might dropout using academic performance data from their first semester. Further, the approach is extended by comparing the results of the cross-program model on specific programs of study with the results of the models trained for each specific program. The findings support the generalization of the approach of Berens et al. (2019) to the German context, which could serve to establish best practices for dropout prediction in higher education.

Keywords: dropout prediction, replication study, machine learning, models' comparison

1 INTRODUCTION AND RELATED WORK

A goal of the European Commission in 2010 was to increase the number of 30-34-year-olds with higher educational attainment from 31% to at least 40% in 2020 (European Commission, 2010). A way to achieve this is to reduce dropout rates. This requires early knowledge of students who run the risk of not completing their studies. The subsequent supervision of first-year students and the support provided by targeted study counseling must be supported by an effective dropout forecast and the analysis of possible causes (Dekker, Pechenizkiy, & Vleeshouwers, 2009).

Looking for a general approach based only on academic performance data, this work replicates to a large extent the work of Berens, Schneider, Görtz, Oster, and Burghoff (2019) – hereinafter also referred to as original study – regarding the chosen algorithms and academic performance features. The purpose is to use the data the university has on the academic performances of their current and former students to identify students at risk at the end of their first bachelor semester. As an extension of the original study, two additional modeling ways are compared: 1) a cross-program model built with the data of three bachelor programs together and evaluated separately on each specific program and 2) three models built for each specific degree program. The comparison of the results of the original study with all our results indicates that the approach of Berens et al. (2019) is generalizable to other higher education institutions in Germany.

Predicting dropout with machine learning algorithms in higher education institutions is an important task and has been investigated in many works (Ochoa & Merceron, 2018). Researchers use sociodemographic data, performance data or a mixture of both to solve this task. Sociodemographic data might include gender, ethnicity, income, date of birth. Performance data might include pre-university grade, major or degree program declared, enrollment in university courses, university grades. Dekker et al.

(2009), Aulck, Nambi, Velagapudi, Blumenstock, and West (2019) or Berens et al. (2019), as the original study, have obtained good prediction results with performance data only; adding sociodemographic data hardly improves the results. The work of Berens et al. (2019) is particularly relevant to us as it predicts students' dropout in a German context which is also the case of the present investigation. Building on these findings, this work uses only performance data. Different kinds of features can be extracted from performance data, in particular, global features and local features as introduced by Manrique, Nunes, Marino, Casanova, and Nurmikko-Fuller (2019). Local features are specific to a particular program of study like grades in the courses of this program. By contrast, global features can be extracted for any program of study like the number of passed exams, the average grade in passed exams and so on. Note that models built with machine learning algorithms that use local features can be trained only with the data of that particular program while models that use global features only, can be trained with data coming from all programs of an institution; we call these last models cross-program models. Dekker et al. (2009) have investigated one degree program only and use local features while Aulck et al. (2019) and Berens et al. (2019) have used global features and build one cross-program model. Manrique et al. (2019) have investigated two programs of study; they have built two models using local features and a cross-program model. Interestingly, the performance of the two models using local features tends to be better than the performance of the cross-program model. This finding leads us not only to replicate the work of Berens et al. (2019) but also to extend it by investigating whether individual models built separately for each degree program using global features give better results than the cross-program model. From a machine learning perspective, more training data is better. This would speak for a model integrating data from different programs of study. However, data from another program of study could also add noise.

No known algorithm works better in all contexts. Dekker et al. (2009) have obtained the best results with decision trees and Aulck et al. (2019) with logistic regression. Three algorithms – logistic regression, random forests, and neural networks – have given very similar results in the work of Berens et al. (2018); the addition of the ensemble method AdaBoost slightly improved the results.

Models are evaluated differently. Dekker et al. (2009), Aulck et al. (2019) and Manrique et al. (2019) have used k-fold cross-validation. Berens et al. (2019) have picked out a single cohort to evaluate their model. This work has used a time-aware validation in the spirit of Krauss, Merceron, and Arbanowski (2019) to reflect the intended use of the model: it is to build with data of passed students to predict whether new-comers might drop out. This approach is also used in Asif, Merceron, Ali, and Haider (2017) or Baneres, Rodriguez, and Serra (2019).

2 METHOD

This study uses data from six-semester bachelor's degree programs, which include 4,312 students from 2005 until summer 2019. The original study takes the data of two German universities: the entire bachelor courses of a state university (SU) with 14,496 records and a private university of applied science (PUAS) with 7,600 records while our work is based on three bachelor programs of a German state university of applied sciences. Our records include for each student the enrollment date in the degree, every single course they enrolled in, the respective enrollment semester and the grade earned, the graduation date and the result for students who completed the degree.

For the further processing of the data some preprocessing was necessary: to take account of changes to the curricula over the years, all data had to be converted to the present curricula and pseudonymization of the records were carried out by aggregation of grade from grades ({1.0, 1.3}, {1.7}, {2.0, 2.3, 2.7}, {3.0, 3.3, 3.7}, {4.0}) to (1.3, 1.7, 2.3, 3.3, 4.0) with 1.3 is the best and 4.0 the worst. Two other possible outcomes of an exam are “not participated” and “failed”. To earn a degree, a student has to successfully pass every single course and has three attempts to do so. To attempt an exam, a student has to enroll in the corresponding course.

Our use-case is to predict students who drop out of the degree. A student switching from one degree to another degree within the same university or to another university is therefore considered as a dropout as well. To uncover students still enrolled in the university but not enrolled in any course of a specific program, we find dropouts from the data: a student that has not enrolled in any course of the degree during more than two consecutive semesters has dropped out of the degree. This threshold results from the longest interruption that we have identified in graduates.

A preliminary exploration of the data on a smaller dataset has shown that students dropping out and students completing the degree strongly differ in the courses of the first semester. The frequency of occurrence of “not participated” and “failed” is much higher for students who drop out than for students who complete their studies. This observation suggests that the use of appropriate algorithms on the data of the courses of the first semester can predict students’ dropout with good results. That’s why we focus on the first semester in this study.

The global features chosen for this study are given in Table 1 right. Our work differs from the original regarding the features in the following points: 1) we don’t distinguish between important and other successfully completed exams because all courses of our first semester are mandatory and considered as core courses of the programs, 2) we don’t distinguish between exams not participated in and no-show exams because this distinction does not exist in our data.

Table 1: Academic performance features – Comparison with Berens et al. (2019)

Berens et al.		This work	
Variable	Values	Variable	Values
No. of important successfully completed exams	1 to 9	No. of successfully completed exams	0 to max
No. of other successfully completed exams	0 to max		
Average grade per semester	1.00 to 4.00	Average grade per semester	1.3 to 4.0
No. of failed exams per semester	0 to max	No. of failed exams per semester	0 to max
No. of exams per semester not participated in	0 to max	No. of exams per semester not participated in	0 to max
No. of no-show exams per semester	0 to max		
Class label	1 = dropout 0 = graduate	Class label	1 = dropout 0 = graduate

In this study we have used the five different algorithms: decision tree, logistic regression, neural network, random forest, and AdaBoost; the implementation was done in the Python scikit-learn library.

Concerning the algorithms in the original study, the replication differs in three points: we have 1) added decision trees due to their good interpretability like logistic regression, 2) used random forest instead of bagged random forest, because the benefit of the bagged version was not clear to us, and 3) added the decision tree to AdaBoost due to good results as a single classifier. The metrics used to evaluate the models are precision, recall, accuracy and area under the ROC curve. The original study uses a classification threshold, which cannot be repeated in the present research because of the time-aware evaluation. Instead, this work has optimized the hyper-parameters of each model by 10-fold cross-validated grid search tuned for the recall metric.

Figure 1 shows the different training/test sets we have used and their numbers of records: [a] cross-program model: the dataset of the three degree programs is split into 80% training data corresponding to students with the oldest matriculation date and 20% test data (students with the newest matriculation date), [b] program-specific models: is similarly split, but for each degree program because we have trained program-specific models, and [c] cross-program model with program-specific test: the training set is the union of the training sets of [b] and the test sets are the same as in [b] (training set [a] is not necessarily disjunct from test set I, test set II and test set III). Variant [a] corresponds to the replication study while variants [b] and [c] are carried out for its extension.

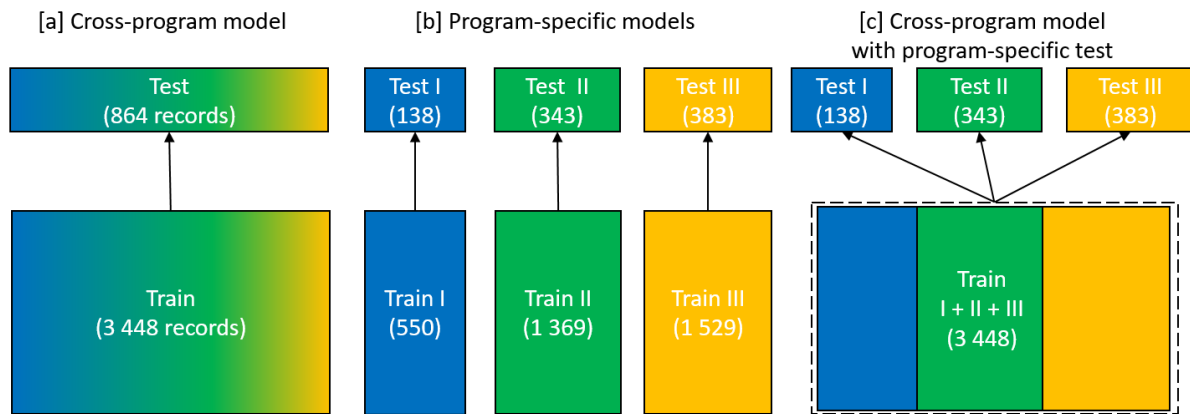


Figure 1: Schematic illustration of the training and test splits

3 RESULTS

Figure 2 presents the scores for each model and each variant [a/b/c]. First, we compare the results of the variant [a] to the results of the original study: recall 71.49% (SU) and 69.89 (PUAS); accuracy 76.60% (SU) and 83.64% (PUAS) – best results obtained with AdaBoost. Our AdaBoost achieves better results: recall 78.19% and accuracy 83.55%; only accuracy (PUAS) is marginally better. AdaBoost outperforms slightly the other models in the original study, which is not the case in our replication. Our other models for variant [a] achieve similar results: recall between 74.49% and 78.56% and accuracy between 81.46% and 83.89%. The best model for variant [a] in terms of recall is the decision tree (78.56%), closely followed by the neural network (78.37%) and in terms of accuracy the neural network (83.89%), closely followed by the decision tree (83.78%). The most important decision feature of the decision tree is the number of successfully completed exams and this is also confirmed by the result of the logistic regression coefficients. It stresses the observation that students who are not successful in their first semester tend to drop out faster.

The comparison of the cross-program models [a/c] with the program-specific models [b] as the extension of the replication study does not show a clear picture. In some cases, the cross-program model outperforms the specific models, for example, the decision tree and degree program III with a recall for [a] of 78.56%, for [b] of 70.64% and for [c] of 78.44%. In other cases, the specific model outperforms the cross-program models like for the AdaBoost and degree program II: recall [b] = 83.98% opposed to recall [a] = 78.19% and recall [c] = 79.84%.

The ROC AUC reaches the highest value for variant [a] with the logistic regression (92.21%). The best AUC score of 95.35% is achieved by the logistic regression in the specific model [b] for program II. AdaBoost performs worse for this metric. Similar trends can be observed for precision.

The differences between the results of the different models tend to be marginal, although the performance for program II tends to be better than for the other programs. Overall, the results show the appropriateness of a cross-program model and confirm the approach of Berens et al. (2019).

Models / Programs		Precision			Recall			Accuracy			ROC AUC		
		a)	b)	c)	a)	b)	c)	a)	b)	c)	a)	b)	c)
AdaBoost	I		93.22%	98.11%		70.51%	66.67%		80.43%	80.43%		81.92%	82.50%
	II	94.63%	96.23%	97.49%	78.19%	83.95%	79.84%	83.55%	86.30%	84.26%	85.37%	87.98%	87.42%
	III		88.78%	84.53%		79.82%	79.36%		82.77%	83.81%		83.24%	84.53%
Decision Tree	I		92.73%	98.15%		65.38%	67.95%		77.54%	81.16%		76.60%	91.42%
	II	94.65%	98.01%	97.06%	78.56%	81.07%	81.48%	83.78%	85.42%	85.13%	90.93%	94.29%	93.58%
	III		89.02%	90.00%		70.64%	78.44%		78.33%	82.77%		86.80%	88.20%
Logistic Regression	I		100.00%	100.00%		67.95%	64.10%		81.88%	79.71%		93.23%	93.93%
	II	95.80%	98.02%	97.47%	75.97%	81.48%	79.42%	82.85%	85.71%	83.97%	92.21%	95.35%	95.02%
	III		95.18%	91.76%		72.48%	76.61%		82.25%	82.77%		89.56%	88.92%
Neural Network	I		98.21%	100.00%		70.51%	67.95%		82.61%	81.88%		84.42%	83.97%
	II	95.07%	97.55%	97.51%	78.37%	81.89%	80.66%	83.89%	85.71%	84.84%	85.77%	88.45%	87.83%
	III		92.47%	91.49%		78.90%	78.90%		84.33%	83.81%		85.21%	84.60%
Random Forest	I		100.00%	98.11%		61.54%	66.67%		78.26%	80.43%		90.49%	91.99%
	II	94.82%	96.53%	97.47%	74.49%	80.25%	79.42%	81.46%	83.97%	83.97%	90.95%	91.98%	93.35%
	III		87.28%	91.26%		69.27%	76.61%		76.76%	82.51%		83.66%	89.17%

Figure 2: Heatmap of achieved metric scores – highest values are in bold

4 CONCLUSION AND FUTURE WORK

In this study, we have investigated five different algorithms using a global feature set to predict students' dropouts using only data from the first semester. Further, we have built cross-program models, models specific to each of the three programs and tested a cross-program model on each study program separately. Overall, the results show that a cross-program model as proposed in the original study is generalizable. Further research is needed to understand why prediction tends to work better for the study program II.

Despite the differences from the original study, our models get comparable results and even better for recall. AdaBoost works best in the original study, which is not the case here. Further investigation is needed to understand why. The obvious next step concerns the dataset: we have used only three degree programs with 4,312 records. So, subsequent activity is to consider more study programs up until a university-wide analysis and prediction system as in Berens et al. (2019). Especially interesting could also be the consideration of different online degree programs as well as the inclusion of master's degree programs.

Further work also consists of taking higher semesters into account. This data might reveal the self-regulating skills of the students better. A preliminary study has shown that about 1/3 of the students who drop out do so during or immediately after the first semester. This means 2/3 of the students drop out later. A follow-up is to predict dropouts related to the semester as in Berens et al. (2019). We will consider how adding more data from higher semesters will impact the performance of the classifiers.

ACKNOWLEDGEMENT

We thank Lennart Egbers and Stephan Wagner for the substantial data preprocessing and their preliminary study.

REFERENCES

- Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. (2017). Analyzing undergraduate students' performance using educational data mining. *Computers & Education*, 113, 177–194. <https://doi.org/10.1016/j.compedu.2017.05.007>
- Aulck, L. S., Nambi, D., Velagapudi, N., Blumenstock, J. E., & West, J. (2019). Mining University Registrar Records to Predict First-Year Undergraduate Attrition. *EDM*.
- Baneres, D., Rodriguez, M. E., & Serra, M. (2019). An Early Feedback Prediction System for Learners At-Risk Within a First-Year Higher Education Course. *IEEE Transactions on Learning Technologies*, 12(2), 249–263. <https://doi.org/10.1109/TLT.2019.2912167>
- Berens, J., Schneider, K., Görtz, S., Oster, S., & Burghoff, J. (2019). *Early Detection of Students at Risk—Predicting Student Dropouts Using Administrative Student Data from German Universities and Machine Learning Methods*. JEDM | Journal of Educational Data Mining, 11(3), 1–41. <https://doi.org/10.5281/zenodo.3594771>
- Dekker, G., Pechenizkiy, M., & Vleeshouwers, J. (2009). Predicting Students Drop Out: A Case Study. *Proceeding of the 2nd International Conference On Educational Data Mining*, 41–50. <http://www.educationaldatamining.org/EDM2009/uploads/proceedings/edm-proceedings-2009.pdf>
- European Commission. (2010, March 3). *Europe 2020—A European strategy for smart, sustainable and inclusive growth*. Retrieved from <https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A52010DC2020>
- Krauss, C., Merceron, A., & Arbanowski, S. (2019). The Timeliness Deviation: A novel Approach to Evaluate Educational Recommender Systems for Closed-Courses. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge - LAK19*, 195–204. <https://doi.org/10.1145/3303772.3303774>
- Manrique, R., Nunes, B. P., Marino, O., Casanova, M. A., & Nurmikko-Fuller, T. (2019). An Analysis of Student Representation, Representative Features and Classification Algorithms to Predict Degree Dropout. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge - LAK19*, 401–410. <https://doi.org/10.1145/3303772.3303800>
- Ochoa, X., & Merceron, A. (2018). Quantitative and Qualitative Analysis of the Learning Analytics and Knowledge Conference 2018. *Journal of Learning Analytics*, 5(3), 154–166. <https://doi.org/10.18608/jla.2018.53.10>

From Data to Intervention: Predicting Students At-Risk in a Higher Education Institution

Irene-Angelica Chounta, Kaire Uiboleht, Kersti Roosimäe, Margus Pedaste, Aune Valk

University of Tartu, Estonia

{chounta, kaire.uiboleht, kersti.roosimae, margus.pedaste, aune.valk}@ut.ee

ABSTRACT: In this paper, we present a computational approach to assess study success in a Higher Education academic institution. To that end, we employ data-mining and machine-learning methods to identify factors that may contribute to students' decision to drop out from their studies and to assess the risk of dropping out for each individual student. In order to communicate the results of the risk assessment, we employ an institutional dashboard – that is, a dashboard that presents the risk assessment per student and the reasons behind this assessment. The institutional dashboard aims to inform academic stakeholders, namely program directors and specialists in academic affairs about reasons that may contribute to dropouts in their programs and to help them identify students that may need further support.

Keywords: Learning Analytics, predictive modelling, dropout, students at risk, learning dashboards.

1 INTRODUCTION

In this paper, we present a research initiative at the University of Tartu in Estonia that aims to employ an evidence-based approach to identify students who may be at risk of dropping out from their studies. As dropouts, we define students' exmatriculations from the respective program for reasons that may reveal students' unwillingness to continue their studies, low academic achievement, lack of motivation or lack of interest. In order to achieve this, we propose a computational approach for assessing students' dropout risk using students' data as recorded by the study information system of the academic institution. The goal is to communicate the results of the risk-assessment through institutional dashboards to academic stakeholders (such as curriculum developers and program directors) so that they can identify bottlenecks in their programs and to provide appropriate feedback and support to students, if needed, in a timely manner.

Securing study success in Higher Education (that is, successful completion of studies leading to an academic degree) is among the goals leading the Europe 2020 strategic agenda¹. Europe aims to scaffold innovation, productivity and also to support social justice by fostering high-level skills through Higher Education. To do that, one of the goals is to increase the rate of young, higher-education graduates by reducing the dropout rates in Higher Education. Estonia has established a number of policies to achieve this goal. However, according to an Annual Report from the Estonian Ministry of Education, the dropout rates for Bachelor students were approximately 51% in 2016² across all

¹http://publications.europa.eu/resource/cellar/d9de3b17-0dcf-11e6-ba9a-01aa75ed71a1.0001.01/DOC_1

² https://www.hm.ee/sites/default/files/annual_analyses_2016_1.docx

disciplines. This finding is supported by related studies showing that dropout rates in Estonian Higher Education Institutions can come up to two thirds depending on the field of study (Kori & Mardob, 2017).

The University of Tartu (UT)³ is Estonia's oldest university and leading centre of research and training. It consists of four faculties: the Faculty of Arts and Humanities, the Faculty of Medicine, the Faculty of Social Sciences and the Faculty of Science and Technology. In 2019 overall, 13400 students – out of which 1660 are international students – study in UT either in the bachelor, master, or PhD programs. In this context, UT launched an initiative in 2019 aiming to support students of mainly Bachelor and Master levels to successfully completing their studies but also to help other academic stakeholders (in this case, program directors and specialists in study affairs) to identify potential reasons that may contribute to dropouts in their programs or curricula and to provide appropriate feedback and support to students.

2 RELATED WORK

Detecting students at risk of dropping out of their studies is a prominent topic of research since dropout rates have a strong impact on the individual (student), the institutional (academic institution) and the national (country) level. Many frameworks have been proposed to evaluate academic success and to identify factors that influence it. For example, Tinto proposed a theoretical model of students' dropouts from college that built on work from social psychology and economics of education (Tinto, 1975, 2017). Tinto's model identifies two dimensions in the model as fundamental for academic success: student's characteristics (such as family background and individual attributes; e.g. goals to study in college) and student's experience with the academic system (such as performance and interactions with teachers and peers). Tinto specified that students need both academic and social integration to ensure retention in studies. According to the model, academic success is affected by the student's individual commitment to their goal along with the student's commitment to the academic system itself.

Arnold and Pistilli (Arnold & Pistilli, 2012) proposed a student success system – Course Signals – in order to support faculty members of a Higher Education Institution (Purdue University) in providing meaningful feedback to students. The system used machine learning algorithms and data mining to predict students who may be at risk of dropping out their studies. The system used data about students' earned credits, student's effort in terms of interaction with the learning environment, students' performance in earlier studies - for example, high school Grade Point Average (GPA) or performance in standardized tests – and other information, such as demographics. Also, Barber and Sharkey (Barber & Sharkey, 2012) proposed the use of predictive models identifying students at risk in the University of Phoenix. In this case, the model combined data from the learning management system, the financial aid system, and the study information system to assess the risk of any given student failing at the course level. Earlier research in Estonian HEIs is based mainly on self-report surveys among dropouts (Kori et al., 2016; Must et al., 2015). The studies show that student dropout is often related to the combination of reasons that include individual and curriculum-level factors: for example, dissatisfaction with the quality or organization of studies, inefficient academic and social

³ <https://www.ut.ee/en/university>

environment, wrong choice of studies, inefficient study skills and low motivation, working during the studies and financial reasons. In this work, we use the findings of related research in context, to predict students' dropout and to inform our research with lessons learned and successful practices from similar studies.

3 METHODOLOGY

The overarching goal of this research is to provide a holistic assessment of students' performance (Chounta et al., 2019) in three ways:

- by using various kinds of data, for example information on the course level from the learning management system the university uses and also from the students' feedback questionnaires;
- by applying multilevel analytical approaches – for example social network analysis and pattern mining – to analyze various data sources and to complement insights; and
- by supporting stakeholders through learning analytics dashboards that will present multimodal feedback, for example textual feedback along with visualizations.

Currently, in order to assess risk of dropping out, we use students' data as recorded in the study information system of the academic institution (University of Tartu). Our dataset includes information about students' demographics, their prior academic background and their progress while studying at the institution. After consulting with the university's academic commission about potential issues regarding privacy and ethics, we decided to exclude demographical information – such as gender or citizenship – or potentially private or sensitive information – such as postal address – when assessing whether a student is likely to drop out or not. To take into account differences between student populations that can be attributed to the curricula or the faculties, we modelled these factors as random effects. For the purpose of this work, we employed a computational model that predicts risk on three dimensions:

- a) *academic background*. That is, information that may relate to student's previous academic experience, such as: admission grade, number of degrees that the student has acquired and how many times a student has been enrolled in the university's study programs;
- b) *effort in terms of participation*. To assess effort, we used the following features: the amount of registered courses and credits, the amount of credits the student cancelled, the amount of credits registered for extra-curricular courses, the time a student spent on academic leave, the time a student spent studying abroad and the student's workload (full or part time);
- c) *performance in terms of academic achievement*. To assess performance, we used the following features: the number of successfully completed courses, the number of failed courses, the number of no-showups in exams, the amount of earned credits and the differentiated scores (for example, amount of A's, number of B's, and so on).

For each of these three dimensions, the computational model – in this case a logistic regression classifier – provides a binary assessment, that is whether the student is likely to dropout or not. We decide on the "severity" of the risk assessment based on the following rule:

- Students who are predicted to drop out on three dimensions are classified as "high-risk";
- Students who are predicted to drop out on at least one dimension are classified as "medium-risk";
- Students who are not predicted to drop out on any dimension, are classified as "low-risk".

The results of this assessment will be presented to program directors through an institutional dashboard. This process is described in Figure 1. With the term “institutional dashboard”, we mean an online interactive and dynamic interface that will be accessible through the study information system of the institution to program directors. The rationale is to inform them so that they can assess potential risks for their respective program, to help them redesign their program if necessary and to support them in identifying specific cases where an intervention might be needed. Neither teachers nor students will have access to the information presented in the institutional dashboard. The reason is that we do not want to create or support any bias either on the student or the teacher level and to potentially affect student’s motivation negatively.

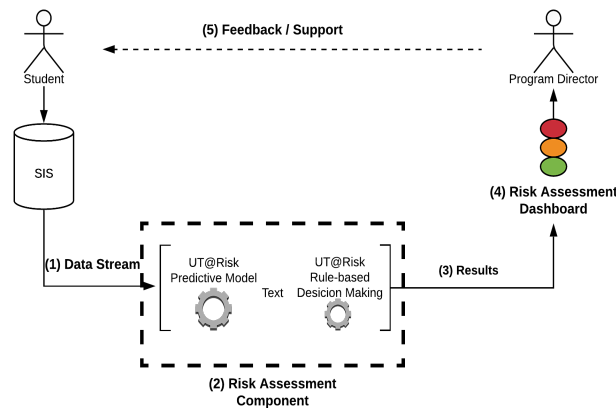


Figure 1. Risk Assessment process for detecting students at-risk of dropping out.

4 FIRST INSIGHTS

In order to test our approach, we collected data from bachelor students and students in Bachelor’s and Masters integrated programme (in Medical Faculty) who enrolled in the university from 2010 to 2014. The rationale was to use data of students whose nominal time of studies (3 years in case of Bachelor programme and 6 years in case of integrated studies) is as a rule over and most of whom should have had the opportunity to graduate. Overall, the dataset contained 3695 students from all four faculties. The distribution and dropout rates of students among the four faculties of the university are presented in Table 1.

Table 1. Number of students and dropouts per faculty

Faculty	Number of Students	Number of Dropouts
Arts and Humanities	599	296 (49%)
Medicine	786	162 (21%)
Social Sciences	1443	582 (40%)
Science and Technology	853	429 (50%)

To train and test the model, we split the dataset into two parts: the training and the test sets. For the training set, we used data of students who were matriculated from 2010 to 2013. For the test set, we used data of students who were matriculated in 2014. This resulted in a 70/30 split: 70% of the original dataset was used for training the model and the remaining 30% was used for testing the model. This decision was made in order to test whether using old data to predict dropouts for recent cases could provide accurate predictions. However, we acknowledge that this can have a negative impact on the accuracy of predictions, especially if the dropout rates have changed significantly over the years. We plan to explore the effect of changes in dropout rates on predictive accuracy in future work.

Table 2. Classification metrics for predictions based on the three independent classifiers and on their combination.

	Performance Classifier	Effort Classifier	Academic Background Classifier	Perf + Eff Classifier	Combined Prediction (RAC)
Recall	0.95	0.90	0.40	0.95	0.97
Precision	0.93	0.94	0.55	0.95	0.95
Accuracy	0.95	0.93	0.60	0.96	0.96
F-measure	0.94	0.92	0.46	0.95	0.96

We tested the performance of the Risk Assessment Component (RAC) on the test set. Overall, the test set consisted of 1248 students out of whom 544 students had dropped out from their studies. The RAC assessed that 514 students were on a high risk of dropping out their studies, 217 students were evaluated as medium-risk of dropping out while 517 students were assessed as low-risk. Out of the 514 students that were predicted as high-risk, 488 students indeed dropped out (96%). Similarly, out of the 217 students who were predicted as medium-risk, 39 students dropped out eventually (20%). Finally, out of the 517 students who were assessed as low-risk, only 17 of them eventually dropped out (3.3%). Table 2 shows the results per independent classifier and for their combinations. The results of the Performance and the Effort Classifier are highly correlated ($p=0.92$, $p<0.001$) while the correlations between these classifiers and the Academic Background Classifier are low ($p<0.2$, $p<0.001$). Nonetheless, including the Academic Background dimension in the classification process appears to provide the best results in terms of precision, recall and accuracy.

5 DISCUSSION

To communicate the risk assessments to the stakeholders, we designed an institutional dashboard following the traffic lights metaphor (Figure 1). That is, students who were assessed as high-risk, were followed by a red traffic light, students who were assessed as medium-risk were followed by a yellow traffic light, and students who were assessed as low-risk, were followed by a green traffic light. This design was presented as a mock-up to approximately 30 program directors from all faculties during one of their regular meetings and it was well-received. Additionally, the program directors indicated that they would like to receive information about the reasoning behind the model's predictions. That is, why the model predicted that a student belongs to a specific risk group. They also commented that it is important for them to receive this assessment in a timely manner – that is, in the beginning of the new semester, so that they have enough time to intervene, if needed.

Personal ID	Student Place ID	Risk Assessment	Risk Assessment Reasoning
0	68807	386242	• no experience studying abroad
1	135475	386248	• big experience as a UT student
2	154127	386249	• no experience studying abroad
3	174565	386255	• few extracurricular credits earned
4	174772	386257	• many days on academic leave
5	191951	386265	• no experience studying abroad
6	191952	386266	• few extracurricular credits earned
7	200570	386272	• little financial support
8	217350	386284	• no experience studying abroad
9	221741	386291	• many days on academic leave
10	221742	386292	• no experience studying abroad
11	236344	386304	• many days on academic leave
12	239999	386330	• no experience studying abroad
13	240009	386340	• no experience studying abroad
14	240019	386350	• many days on academic leave
15	240020	386351	• no experience studying abroad
16	240024	386355	• little financial support
17	240028	386359	• many days on academic leave
18	240049	387034	• no experience studying abroad

Figure 2. Institutional Dashboard Mockup

Currently, we are re-designing the institutional dashboard and the predictive approach taking into consideration stakeholders' feedback and adding functionality, such as historical data about dropouts in the respective curriculum and current trends. Additionally, we plan to carry out extensive design workshops and pilots with the participation of program directors and curriculum developers as well as stakeholders from the university's government and administration. In this way we want to ensure that the institutional dashboard reflects existing needs and standards of the academic community and that the dashboard's interface is usable and useful for the target user population.

This paper presents work in progress and we acknowledge that significant improvements will be required until we reach the state of launching a viable solution. At the same time, we are aware of existing limitations. As aforementioned, program directors requested timely assessments – the earlier, the better. This is a challenging task especially for first-year students since we rely mostly on metrics of academic effort and performance and we do not take into account students' demographics. One potential solution would be to use student-entered data about their goals and expectations regarding the institution and their motivation for pursuing an academic degree. Another limitation is the way risk assessments should be used. At this point, we do not plan to use this information in any other way rather than for reflecting on our practices and policies. For example, to reflect on what measures could the university take to support students in pursuing their degree. Even though some stakeholders voiced their willingness to intervene with their own means in critical cases, this bears the questions: what would be an appropriate intervention taking into account that students in Higher Education are adults and what would be the cost of it?

REFERENCES

- Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 267–270.
- Barber, R., & Sharkey, M. (2012). Course correction: Using analytics to predict course success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 259–262.
- Chounta, I.-A., Pedaste, M., & Saks, K. (2019). Behind the Scenes: Designing a Learning Analytics Platform for Higher Education. *Companion Proceedings 9th International Conference on Learning Analytics & Knowledge (LAK19)*. International Conference on Learning Analytics & Knowledge.
- Kori, K., & Mardob, K. (2017). *First-year higher education ICT studies and dropout rates—the Estonian case*.
- Kori, K., Pedaste, M., Altin, H., Tõnisson, E., & Palts, T. (2016). Factors That Influence Students' Motivation to Start and to Continue Studying Information Technology in Estonia. *IEEE Transactions on Education*, 59(4), 255–262.
- Must, O., Must, A., & Täht, K. (2015). Programmi TULE uuringu „Haridustee vali kud ning õpingute katkestamise asjaolud Eesti kõrghariduses“ aruanne. *Psühholoogia Instituut. Tartu Ülikool*.
- Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, 45(1), 89–125.
- Tinto, V. (2017). Through the eyes of students. *Journal of College Student Retention: Research, Theory & Practice*, 19(3), 254–269.

For and by Student Dashboards Design to Address Dropout

Benjamin Gras, Armelle Brun and Anne Boyer

Université de Lorraine, LORIA - FRANCE

{Benjamin.gras, armelle.brun, anne.boyer}@loria.fr

ABSTRACT: This paper focuses on the design process of a student dashboard, in the frame of a Learning Analytics project. The dashboard is intended to reduce dropout of first-year University students. The strong points of this dashboard are three-fold: 1) the involvement of students in the whole design process, 2) the possibility of students to personalize the dashboard, 3) the possibility of indicating personalized goals.

Keywords: Student dashboard, student dropout, higher education.

1 INTRODUCTION

The French project EOLE (Engagement to Open Education - <http://www.dune-eole.fr>) aims at designing a different approach of education at the University, both in its modalities and in the enlargement of its target audience. This major challenge is fundamental to develop the contribution of universities in the learning sector, in both a citizen and a competitive approach. One of the numerous goals in EOLE is to address students' dropout problem, mainly for first-year students, who are known to massively dropout.

Two achieve this goal, EOLE relies on two main hypotheses: 1) most of the students who dropout lack information about their learning behaviour: the way they learn, compared to others or not, about their progress related to the expected output, etc. So, if students can get more information, dropout will decrease; and 2) if students can feel under control of their learning process or feel heard, by informing their expectations, receiving advice, etc., dropout will also decrease.

Based on these hypotheses, EOLE proposes students to access course-level personalised dashboards. To ensure a high quality dashboard, a multi-profile team has been set up: teachers, students, vice-rectors, computer scientists, researchers, among others. For the sake of representativeness, teachers and students from a great diversity of disciplines are involved in the team. This multi-profile team and the involvement of students in all steps of the project are one strong point of this project.

The dashboard, and the associated features and indicators, are further presented and explained below. A study of the impact of the dashboard on student dropout will be conducted in the following weeks.

2 RELATED WORKS

In the literature, self-regulated learning (SRL) can be defined as being "an active, constructive process, where learners define their learning objectives and try to supervise, regulate and control their cognition, motivation behaviours, guided and constrained by their objectives and

characteristics related to the environment" (Pintrich, 2000). Zimmerman (Zimmerman, 2002), also explains that the differences in learning success are mostly attributed to the self-regulation ability of learning, which are relevant to the initiation and maintenance of the learning process. In addition, a recent study by Aljohani (Aljohani et al., 2019) shows that student-centred dashboards (Govaerts et al., 2012, Odriozola et al., 2012) increase student engagement (investment in time, etc.) more than teacher-centred dashboards (Guo et al., 2017) (in the latter case student engagement could be increased through the interaction between students and teachers). In this latter study, students can consult a dashboard giving them statistical, graphical and textual feedback about their learning. The use of this dashboard by students has been tracked and an analysis shows that students who use the dashboard are significantly more engaged (i.e. spend more time on the platform and have more activities on the forums).

3 BUILDING A STUDENT-CENTERED DASHBOARD

As highlighted in the literature, dashboards are a way to support students in the self-regulation of their learning. EOLE assumes that it can also be a way to address student dropout, and thus proposes to design a dashboard that is targeting students. The key point in this design is that students are at its core.

In a preliminary phase, students have been invited to share their needs, in terms of features/functionalities of a dashboard. A needs analysis has been conducted with about 100 first-year students. Below are the most recurrent needs that students expressed. Notice that some of them have been highlighted in a similar study (Schumacher & Ifenthaler, 2018).

- Indicators should be sufficiently diversified so that every student can find those corresponding to his/her wishes.
- Obsession with indicators should be avoided.
- Indicators must be beneficial and their reading must be easy.
- Indicators should value the advice between peers. Senior students should volunteer to mentor junior students.
- Advice about the methodology of academic work (organization, work methods) is welcome, not just help about course content.

This needs analysis resulted in the design of a first prototype of the student dashboard. It has then been presented to other first-year students to obtain their feedback about the features and the indicators proposed. In all, more than 300 first-year students, spread over several iterations, gave their opinion during the iterative and incremental co-design of the dashboard.

A strong assumption on which EOLE relies is that a dashboard is a tool that should be made for the students' own interest, and that it should not be intended to constrain students. To ensure that, at each iteration of the dashboard design, students' opinions were collected through a questionnaire. The first version of the dashboard (from the first iteration) was presented to 88 students, along with questions about the features they think they would use if these features were made available to them. The results of this questionnaire are presented in Table 1. Although the literature highlights the comparison with peers, especially in higher education, only 56% of the students are in favour of this feature, i.e. nearly half of the students do not wish to compare themselves with their classmates. In addition, 5 students (6%) expressed their fears and apprehensions about the impact of comparing themselves with peers on their personal well-being. It has thus been decided to display the peer comparison feature only on demand of the student. Thus, students who want this feature

have to explicitly tick the appropriate answer. Other less requested features were not explicitly criticized by students, so it has been decided to keep them on the dashboard.

Table 1: What feature students wishes in a dashboard?

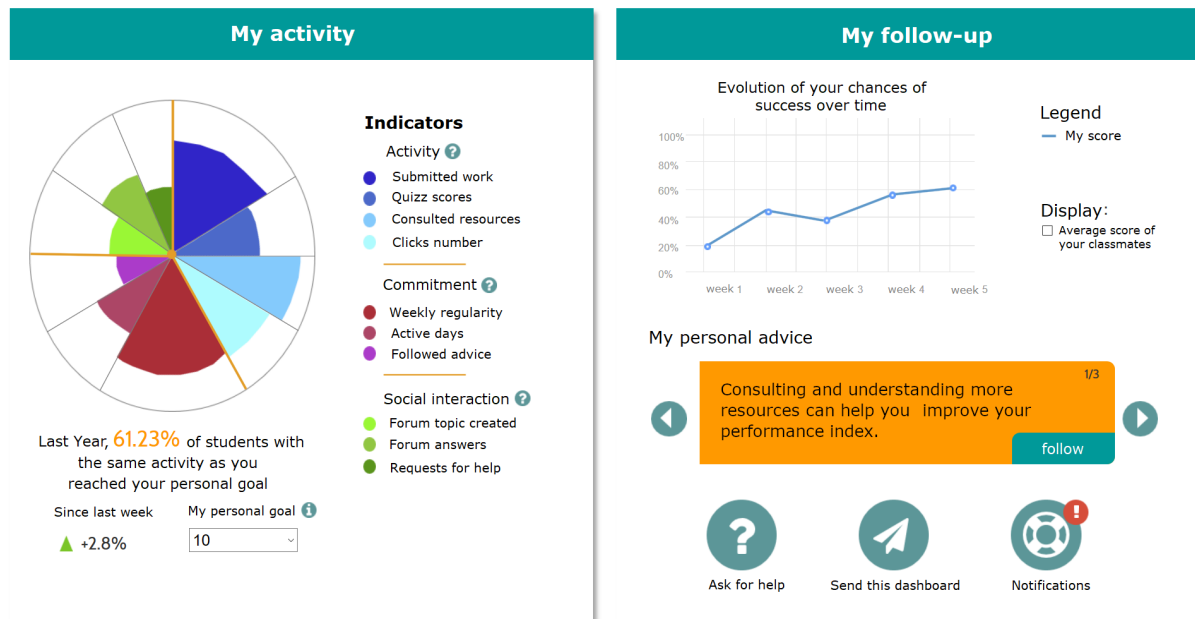
Feature	Percentage of students who would use the feature (out of 88)
Individual performance	99%
Peers comparison	56%
Automatic advice	52%
Help other students	48%
Ask for advice	38%

The final version of the dashboard (presented in Figure 1) has been obtained after three iterations. The dashboard displays indicators about the activity of a specific student in his/her Algorithms and programming course. It is divided into two parts. This final version is available at the following [link](#). Since this is an interactive and customizable dashboard, the best way to understand it is to interact with it directly online.

3.1 “My activity” part of the dashboard

The left part of the dashboard, named “My activity” (Figure 1) displays raw indicators of the student’s activity. This part includes the most awaited functionality by the students: the individual performance (99% of students have declared to want it). Three types of raw indicators have been selected:

- Activity indicators (blue ones): number of submitted works, quiz scores, number of resources viewed and the total number of actions on the course.
- Student engagement indicators (red ones): number of active days, weekly regularity and number of completed automatic advice. The weekly regularity of a student has been adapted from (Boroujeni, Sharma, Kidziński, Lucignano, & Dillenbourg, 2016), which shows not only that students easily understand the meaning of this indicator but also that they are interested in discovering if they are working less than in previous weeks. A correlation of 0.28 is observed between the weekly regularity that we have adapted and the course score. This correlation is calculated from the students’ traces of activity enrolled in the course during the previous years and the final results of the students. With a p -value of 0.002, we can conclude that there is a significant link between the weekly regularity of student work and their academic performance.
- Collaboration indicators (green ones): number of created topics on the forum, number of answers, number of times the student asked for help.

Course: Algorithms and programming**Figure 1: Final version of the dashboard**

Besides, during the co-design iterations, most of the students have put forward the fact that learning traces collected by the system only represent a partial view of their activity. Based on this feedback, it has been decided to add an edit function to allow students to modify the indicators displayed on the dashboard. Thus, these user-modified indicators should better reflect students' actual learning activity. More importantly, students stay in control of their personalised dashboard.

Last, an additional indicator requested by students and proposed in this dashboard, is the student's overall performance (lower section of the left part). To make this indicator possible, students are proposed to provide, on the dashboard, the score that they would appreciate to achieve on the final exam (the personal goal). In Figure 1, the expected score given by the student is 10 (out of 20). In this case, the associated overall performance indicator is 61.23%. This indicator is evaluated as the odds percentage that a student achieves his/her personal goal. As students directly set their personal goal, the student's overall performance is directly influenced. Let two students have the same value on two indicators, the one who fixed his/her personal goal to 18 will not have the same odds percentage as a student who fixed his/her personal goal to 10.

Many indicators that are proposed in the dashboard, such as the indicator of future success, which depends on students' personal goal, but also the fact that students can modify the content of the indicators, have not only been designed to increase the information students can access, but also to increase their feeling of being heard and understood. The expected effect is a decrease of student dropout, especially for students who may lack of confidence in themselves.

3.2 "My follow-up" part of the dashboard

The right part of the dashboard (Figure 1) is divided into three elements.

- At the top, the evolution of the student's performance over time is displayed, in the form of a line chart. This is where the student can choose to display the average performance of his/her classmates.

- In the middle of this part, personalized advice is provided to the student to help him/her improve his/her performance (orange rectangle). The student can follow or not the advice, depending on his/her goodwill.
- At the bottom, two action buttons allow the student to ask for help. The first one is dedicated to receiving help from the teacher. By clicking on this button, the student also accepts to share the data displayed in his/her dashboard. The second one, labelled “send this dashboard”, only shares a capture at time t of the dashboard. The last button is a notification queue to manage the actions of the first two buttons.

The highly personalisable side of this dashboard is intended to make students feel understood and may access the dashboard more often, which could reduce dropout, as students may feel less isolated.

4 USABILITY

The usability of the dashboard proposed here has been evaluated with the System Usability Scale (SUS) (Bangor, Kortum, & Miller, 2008). Although this scale does not allow to strictly quantify the usability, the score obtained (between 0 and 100) allows to locate the perceived usability of the dashboard by the student. 127 students took this well-known test of the user experience literature. The results obtained are presented in Figure 2.

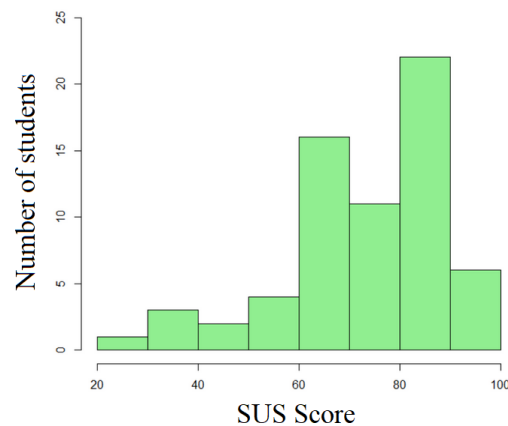


Figure 2: SUS results repartition

We observe the 1st quartile at 65, the median at 75 and the third quartile at 85. The average score given is 74.12, the minimum 27.5 and the maximum 100. In UX Design methods (Lallemant & Gronier, 2015), the authors propose an interpretation scale of the SUS score. Figure 3 presents the associated interpretation scale.

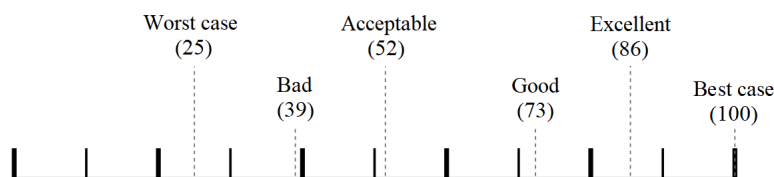


Figure 3: Interpretation scale of SUS score

With an average score of 74.12, the dashboard proposed is between "Good" (73) and Excellent (86), which is rather promising for our imminent 1st live study.

5 CONCLUSION

The work presented in this paper focuses on a student dashboard design process, adopted by the team of the EOLE project. This dashboard is aimed, among others, to decrease student dropout. Its design involves a multi-profile team, including students, who are the recipients of the dashboard. The further step is the test of this dashboard and its actual impact on dropout.

REFERENCES

- Aljohani, N. R., Daud, A., Abbasi, R. A., Alowibdi, J. S., Basher, M., & Aslam, M. A. (2019). An integrated framework for course adapted student learning analytics dashboard. *Computers in Human Behavior*, 92, 679–690.
- Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction*, 24(6), 574–594.
- Boroujeni, M. S., Sharma, K., Kidziński, Lukasz, Lucignano, L., & Dillenbourg, P. (2016). How to quantify student's regularity? *European Conference on Technology Enhanced Learning*, 277–291. Springer.
- Guo, J., Huang, X., & Wang, B. (2017). MyCOS Intelligent Teaching Assistant. 392--393.
- Govaerts, S., Verbert, K., Duval, E., & Pardo, A. (2012). The student activity meter for awareness and self-reflection. *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, 869–884.
- Lallemand, C., & Gronier, G. (2015). *Méthodes de design UX: 30 méthodes fondamentales pour concevoir et évaluer les systèmes interactifs*. Editions Eyrolles.
- Odriozola, S., Luis, J., Verbert, K., & Duval, E. (2012). Empowering students to reflect on their activity with StepUp!: Two case studies with engineering students. *Proceedings of ARETL'12 2nd Workshop on Awareness and Reflection*, 931, 73–86.
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In *Handbook of self-regulation*(pp. 451–502). Elsevier.
- Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior*, 78, 397–407.
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64–70.

Curriculum Analytics as a Communication Mediator among Stakeholders to Enable the Discussion and Inform Decision-making

Liyanachchi Mahesha Harshani De Silva¹, María Jesús Rodríguez-Triana¹, Irene-Angelica Chounta², Kairit Tammets¹, Shashi Kant Shankar¹
Tallinn University¹, University of Tartu²
mahesha@tlu.ee

ABSTRACT: Despite the momentum gained by Competence-Based Learning (CBL), certain crucial aspects such as the competence assessment remain still open. The lack of systematic evidence about how those competencies are delivered and acquired limits different stakeholders -namely curriculum designers, teachers and students- from understanding what is the current state and how to intervene to better support the competence development. To support the stakeholders in this endeavor, this paper presents our approach and plan towards enabling evidence-based decision making in competency-based programs delivered at Tallinn University. More concretely, our proposal collects data from course designs (created by the teachers) and learning traces, and analyses them in the framework of the competence model prescribed at the national level, extracting to what extent competencies have been included in the curriculum and acquired by the students. To reach this goal, we propose a Design-Based Research approach where solutions will be iteratively designed, applied, assessed and refined, involving the different stakeholders in the process.

Keywords: Curriculum Analytics, Competence-Based Learning, Evidence-Based Decision Making, Design-Based Research

1 INTRODUCTION

As raised in multiple International and European reports, there is a global concern about student success and dropout rates in Higher Education Institutions (HEIs) (Vossensteyn et al., 2015). HEIs are trying to address this issue by improving teaching practices and curriculum (Hilliger et al., 2019), e.g., supporting teachers in better selecting, planning and designing suitable activities for the students (Vergas et al. 2019). However, improving the teaching strategy is not enough for student success: improvements at the curriculum level are also necessary (Gottipati & Shankararaman, 2018). For that purpose, Curriculum Analytics (CA) is a systematic approach used by HEIs to develop a curriculum (Hilliger et al., 2019) but still exploratory and mainly oriented towards teachers, leaving curriculum developers' aside (Hilliger et al., 2019).

Among other disciplines, Tallinn University (TLU) has developed CBL study programs in the teacher education degrees to help preservice teachers acquire the competencies defined in national-level teacher qualification standard¹. Thanks to the infrastructure available (namely, eDidaktikum², ad-hoc CBL management system), each subject can be designed attending to the national qualification standards, specifying which competencies are trained in different learning activities. Based on these data, it would be possible to obtain the overall map of how competencies are trained in each degree

¹ <https://www.kutsekoda.ee/>

² <https://edidaktikum.ee/>

as well as extracting which competencies have been already acquired by each student. However, this potential has not yet been explored and there is no systematic evidence about how these competencies from the professional qualification standards are covered in the study programs. Thus, the goal of this study is to support the different stakeholders (mainly, curriculum or program designers, teachers and students) to raise awareness of how the competencies are distributed in the teacher education curriculum, informing potential interventions towards the improvement of teacher education programs, and contributing to the student awareness and decision-making regarding the competence profile they want to achieve.

2 CA FOR CBL: THE TLU CASE SEEN FROM THE SHEILA FRAMEWORK

To prepare HEIs for the integration of learning analytics solutions, the EU project SHEILA created a framework³ to guide institutions during the implementation process. Following this framework, this section provides an overview of our CA initiative towards student success in CBL programs at TLU:

Political context. Aligned with the European guidelines for HEIs (Vossensteyn et al., 2015), TLU is elaborating and implementing strategies to address student success and dropout, e.g., by refining and supporting CBL programs in collaboration with other Estonian universities, TLU has developed eDidaktikum a learning management system that enables teachers to connect their courses to the program and competence model, and also map the competencies to be trained to learning outcomes. Even though the university expected to assess and recognize the competencies achieved by the students at the end of the program, doing it systematically is still extremely challenging. While teachers are expected to use eDidaktikum in their practice, not all of them use it (e.g., in Early Childhood Education, out of 42 courses, 18 are delivered through eDidaktikum, 9 using Moodle, and the rest use other platforms. In addition, the preliminary analysis shows that, in general, teachers are not mapping the competence model with the learning outcomes (e.g., in Early Childhood Education program, only 2 courses have assigned competencies from its corresponding model) and, on the other hand, multiple courses in this program include competencies from other related models).

Stakeholders. The main actors who can contribute to the improvement of CBL programs are: the *program designers*, who are in charge of implementing the national curriculum into the degree; the *teachers*, who connect competencies and learning goals in their courses; and *students*, who should be aware of the competencies that they have or would like to acquire to succeed in their career.

Desired behaviour changes. The main goal of the proposed solution is to raise awareness and support decision making in CBL programs. Program designers will be able to identify the gaps between the competence model defined by the government and the ongoing curriculum offered by the university, triggering potential interventions and contributing to the coordination with the teachers in charge of the different subjects. Teachers will be able to understand the competence level of the students and how their subjects fit in the overall program, refining the course program accordingly. Students will be able to get more global information on their level of competency achievement and based the course selection on the competencies that they want or need to achieve at the end of their studies.

³ <https://sheilaproject.eu/sheila-framework/>

Develop engagement strategy. To achieve that goal, our CA solution will combine systematic tagging and monitoring of competencies in CBL programs. Following the Design-Based Research (DBR) methodology (Wang and Hannafin, 2005), we will iteratively understand the current practices and develop solutions involving the different stakeholders. More concretely, we will gather data about how TLU has implemented CBL programs in eDidaktikum, and compare the competencies expected in those curriculums with the ones developed in the different subjects, and finally acquired by the students. The first study will take place on the first semester of the 2020-2021 course and 6 teachers together with the program designer will participate. Thereafter, the study will expand up to the whole program iteratively (addressing all teachers), other CBL programs and other LMS platforms. Also, in future stages, we will progressively involve students.

Internal capacity to affect change. This initiative is aligned with TLU's goal and initiatives to change the mindset among university members towards evidence-based decision-making, especially promoting the adoption of learning analytics solutions. Regarding ethics and privacy issues, following the EU regulations and the institutional policies, this proposal will be GDPR compliant and will follow the Estonian research code of conduct. In addition, we will involve the TLU GDPR specialist and an institutional expert on ethics to validate and refine the proposal.

Monitoring and learning frameworks. Inspired by the EFLA framework⁴, an evaluation instrument will be developed to assess the performance, effectiveness, impact and maturity of our CA solution with our three stakeholders. The evaluation will be carried out after each iteration of the DBR lifecycle.

3 CONCLUSION

This paper proposes a CA solution and process to support decision-making among the different stakeholders of CBL programs. Given the wide adoption of CBL, we expect to inspire other workshop researchers and practitioners who can promote student's success by improving the curriculum. Similarly, we expect to gather feedback and recommendations to refine the future iterations of our DBR process.

REFERENCES

- Gottipati, S., & Shankararaman, V. (2018). Competency analytics tool: Analyzing curriculum using course competencies. *Education and Information Technologies*, 23(1), 41-60.
- Hilliger, I., Miranda, C., Celis, S., & Pérez-SanAgustín, M. (2019). Evaluating Usage of an Analytics Tool to Support Continuous Curriculum Improvement. EC-TEL 2019
- Vargas, H., Heradio, R., Chacon, J., De La Torre, L., Farias, G., Galan, D., & Dormido, S. (2019). Automated Assessment and Monitoring Support for Competency-Based Courses. *IEEE Access*, 7, 41043-41051.
- Vossensteyn, J. J., Kottmann, A., Jongbloed, B. W. A., Kaiser, F., Cremonini, L., Stensaker, B., ... Wollscheid, S. (2015). Dropout and completion in higher education in Europe: main report. European Union.
- Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational technology research and development*, 53(4), 5-23.

⁴ <http://www.laceproject.eu/evaluation-framework-for-la>

Learning Analytics of Student-Centered Advisory Systems in the Introductory Phase of Teacher Education

Natalie Kiesler

Goethe University Frankfurt
kiesler@em.uni-frankfurt.de

ABSTRACT: The goal of this work is to introduce the online advisory structures (OSA) for teacher education at Goethe University. The recently developed online courses offer information, orientation and student support at the introductory phase of study, particularly addressing Generation Y by using videos, interactive elements and feedback on students' choice of study, study motives and expectations of their future professional practice. Thus, it is the goal to support prospective students in their reflective choice of study and career as teachers. LernBar Studio as authoring tool along with the university's OSA-platform are used to display, publish, distribute and evaluate the produced contents ever since June 2019. User log files and interaction data are saved in a learning record store, constituting the basis for future analyses. At this early stage, not all data has been evaluated. Nonetheless, the recorded data offers insights into students' demands for information during the introductory phase of study. The data also has the potential to improve online counseling services and therefore, address dropout rates in teacher education in the long run.

Keywords: Teacher education, Online Advisory System, Online Counseling, Online Self-Assessment, Self-reflection, Study entry phase

1 THE DEMAND OF ASSESSMENT IN TEACHER EDUCATION

Assessing students prior to university entrance is an area of great interest in teacher education. Teacher students' career choice motives along with certain character traits influence student success at a large scale. It is for this reason that links to study and job satisfaction, resilience, subjective exhaustion and overload were investigated and established. Several studies show that intrinsic motivation leads to a high satisfaction with the study program and finally to student success (Blömeke, 2009; Mayr, 2012; Rothland & Terhart, 2011). Certain traits of character along with career motives cause students to be more satisfied during their studies, thereby decreasing the danger of dropping out. Moreover, they enhance engagement (Rauin & Meier, 2007). Self-regulation competence for instance can help avoid emotional distress and dissatisfaction in the professional life. Students choosing to be a teacher for reasons of compatibility (of family and career) experience a higher amount of work-related stress since their understanding of the duties and tasks is less realistic (Abele 2011, p. 691). However, dropouts result from various factors ranging from individual and institutional to social ones, for instance, study motivation, performance, finances and study conditions at university (Heublein & Wolter 2011, pp. 223-228).

Currently, a reliable pre-selection of candidates for teacher education is not possible. Numbers of applicants at university vary each semester, so does the number of demanded teachers. (Blömeke, 2009; Mayr, 2012; Rothland & Terhart, 2011). Moreover, automated assessment requires a not yet existent, scientifically valid and legal justification. This is way counseling and information related to

occupational aptitude is recommended instead (Rothland & Terhart, 2011; Schaarschmidt, 2012). At an early career stage, counseling can help address issues related to study conditions contributing to dropouts (Heublein & Wolter, 2011). Accordingly, Goethe University in association with the state Hesse set it as a goal to develop 20 online advisory structures, so-called “Online Studienwahl Assistenten” (OSA) until 2020 in order to help improve the transitional phase between school and university (Goethe University, 2016). Therefore, the author designed and implemented four OSA, whereas each OSA represents one in four different teacher study program offered at Goethe University, including teaching at primary schools (L1), secondary schools (L2), grammar schools (L3) and special-needs schools (L5). All four OSA were launched in June 2019. The resulting user data will be analyzed according to whether and how prospective students and freshmen use the system.

2 OSA STRUCTURE, IMPLEMENTATION AND DATA GATHERING

All four OSA for teacher students at Goethe University are available online via the university’s OSA portal. They were implemented by using the SCORM-compatible authoring tool LernBar Studio which allows for the integration of content and media via templates, and several QTI-compatible question formats with feedback options (Voß-Nakkour, 2013). The online units were launched with the support of studiumdigitale who developed advanced feedback options. Participation is voluntary, but users are required to submit a short registration including demographic data and agreement to the data collection declaration according to the General Data Protection Regulation. Upon agreement, users receive information on the university, the structure of the study program, admissions, student life, teaching methods, lectures and extracurricular activities. The information is presented in form of texts, pictures and 15 video-interviews of actual students and academic staff. Second, each OSA offers several different content-related tasks such as open or multiple-choice questions developed by the faculties represented in the study program (computer science, physics, didactics, theology, etc.). Third, individual motives for one’s choice of study are evaluated by comparison with data from a reference group consisting of students having finished their internship at school. The respective items for self-reflection come from a validated instrument (so-called “Fit-Choice”). Thereby well-known study motives of teachers are queried, such as intrinsic motivation, compatibility of family and career, positive influence of third parties, perceived teaching qualification, making a social contribution, helping to shape the future of young people, or the stopgap solution to give some examples. The automatically generated feedback points out how the user responds relative to the peer group by means of a visualization and textual information.

Thus, the system does not only collect demographic data, but also the user’s behavior on the platform, focusing on interactive elements such as questions and the use of videos. The answers to all questions (content-related, self-reflective and evaluative) are saved along with a timestamp. Moreover, the number of clicks on play and pause video buttons, the respective viewing time, completion rate and timestamp are recorded for each viewed video. It is not yet possible to define the overall duration of OSA usage. The gathered data is anonymized, saved in a learning record store and currently sent to the author as Microsoft Excel file which has to be evaluated manually. At this early stage, all data is analyzed by humans, leading to time-consuming quantitative and qualitative analysis of user input, for instance when it comes to the responses to open questions. At this point, students’ motives according to the self-reflective questions require further analysis. The video data has been evaluated and shows high completion rates, the same is true for the OSA in general.

3 DISCUSSION AND CONCLUSION

Academic integration, the first-year experience and student orientation are some of the crucial factors for student success (Heublein & Wolter, 2011). However, investigating student dropout is still an emerging field in teacher education. Stakeholders are particularly interested in the integration of introspection, as teacher's job satisfaction, resilience and performance highly depends on individual motives. Generally speaking, participation in study counselling or individual coaching can help reduce dropout rates (Kot, 2014). Therefore, the development and implementation of the described online advisory structures (OSA) seems a reasonable measure to address dropouts at an early stage. Currently, the entire set of recently gathered data is still being analyzed. In the long run, further in-depth qualitative research, as a part of a long-term study is required to investigate any possible impacts of OSA on dropout rates. In the future, the study of correlations with student demographics, prior knowledge and other factors is intended in order to help understand student success and dropout and develop further adequate interventions for teacher students in the future.

REFERENCES

- Abele, A. E. (2011). Prädiktoren des Berufserfolgs von Lehrkräften. Befunde der Langzeitstudie MATHE. *Zeitschrift für Pädagogik*, 57(5), 674–694.
- Blömeke, S. (2009). Ausbildungs- und Berufserfolg im Lehramtsstudium im Vergleich zum Diplom-Studium - Zur prognostischen Validität kognitiver und psycho-motivationaler Auswahlkriterien. *Zeitschrift für Erziehungswissenschaft*, 12(1), 82–110.
- Goethe University, (2016). Zielvereinbarungen 2016 bis 2020 zwischen der Goethe-Universität Frankfurt und dem Hessischen Ministerium für Wissenschaft und Kunst.
- Heublein, U., & Wolter, A. (2011). Studienabbruch in Deutschland. Definition, Häufigkeit, Ursachen, Maßnahmen. *Zeitschrift für Pädagogik*, 57 (2011) 2, 214-236.
- Kot, F. C. (2014). The impact of centralized advising on first-year academic performance and second-year enrollment behavior. *Research in Higher Education*, 55, 527–563.
- Mayr, J. (2012). Ein Lehramtsstudium beginnen? Ein Lehramtsstudium beginnen lassen? Laufbahnberatung und Bewerberauswahl konstruktiv gestalten. In Weyand, B., Justus, M., & M. Schratz (Eds.), *Auf unsere Lehrerinnen und Lehrer kommt es an* (pp. 38-57). Essen: Stifterverband für die Deutsche Wissenschaft.
- Rauin, U., & Maier, U. (2007). Subjektive Einschätzungen des Kompetenzerwerbs in der Lehramtsausbildung. In M. Lüders, & J. Wissinger (Eds.), *Forschung zur Lehrerbildung. Kompetenzentwicklung und Programmevaluation* (pp. 103-132). Münster: Waxmann.
- Rothland, M., & Terhart, E. (2011). Eignungsabklärung angehender Lehrerinnen und Lehrer. Einführung in den Thementeil. *Zeitschrift für Pädagogik*, 57 (5), 635-638.
- Schaarschmidt, U. (2012). Eignung für den Lehrerberuf frühzeitig erkennen und kontinuierlich fördern. In Weyand, B., Justus, M., & M. Schratz (Eds.), *Auf unsere Lehrerinnen und Lehrer kommt es an. Geeignete Lehrer/-innen gewinnen, (aus-)bilden und fördern* (pp. 58-77). Essen: Stifterverband für die Deutsche Wissenschaft.
- Strang, K. (2016). Beyond engagement analytics: which online mixed-data factors predict student learning outcomes? *Education and Information Technologies*, 1-21.
- Voß-Nakkour, S. (2013). *Lean Media Production: Konzept und Unterstützung durch das Autorensystem LernBar*. Doctoral Dissertation. Goethe-Universität Frankfurt.

A Survey of Learners' Video Viewing Behavior in Blended Learning

Mehrasa Alizadeh¹, Shizuka Shirai², Noriko Takemura³, Shogo Terai⁴, Yuta Nakashima⁵,
Hajime Nagahara⁶, Haruo Takemura⁷

Osaka University, Japan

alizadeh.mehrasa@lab.ime.cmc.osaka-u.ac.jp¹, shirai@ime.cmc.osaka-u.ac.jp²,
takemura@ids.osaka-u.ac.jp³, terai.shogo@lab.ime.cmc.osaka-u.ac.jp⁴,
n-yuta@ids.osaka-u.ac.jp⁵, nagahara@ids.osaka-u.ac.jp⁶, takemura@ime.cmc.osaka-u.ac.jp⁷

ABSTRACT: This study analyzes 19 students' interaction patterns with 6 video lectures in a blended course, using log data and video viewing behavior. We took into account the actions that learners took during the online coursework by checking their screen captures and face recordings to delve more deeply into the nature of their evolving interactions and decisions. The results have revealed the existence of four groups of learners. Our findings provide evidence for the importance of triangulating data sources on learners' video watching behavior to enhance feedback provision to at-risk learners and lower dropout rates.

Keywords: video analytics, learner-video interaction, learning analytics, blended learning.

1 INTRODUCTION

Online learning, in particular the use of videos in higher education has expanded massively over the past 20 years. Much has been written on the benefits of online learning, yet while the popularity of online learning has improved, the issue of high dropout rates has come to light (Tan & Shao, 2015). In order to address this issue, the field of learning analytics has observed a growing interest in extracting log data from learners' use of videos and analyzing them to provide timely support to at-risk learners. Nonetheless, the majority of these studies mainly rely on data from learners' use of videos as indicated by their clickstream and barely move from usage to engagement (Mirriahi & Vigentini, 2017). Using clickstream data has been proven to be a successful method in grasping learners' behavior trends in large-scale MOOC studies. However, in blended learning contexts, we need to gain a deeper understanding of learners' behaviors, for instance wakefulness and motivation, to utilize it for improving face-to-face classes. In order to fulfil this aim, we have collected and analyzed data from students in an introductory blended course on informatics for social-science majors. It is worth emphasizing that this study focuses on specific learning processes through analyzing students' video viewing behavior by monitoring their on-screen actions and checking for their facial expressions and wakefulness while watching video lectures, a type of data that is missing in large-scale MOOC studies. As a result, our findings help understand learner preferences and interaction styles alongside feeding information into the design of an adaptive learning system currently under construction.

METHOD

The study was conducted with 19 first-year undergraduate students (14 females, 7 males, mean age=18.28) enrolled in a blended course titled "Informatics Basics for Social-Science Majors." For the purpose of this analysis, we have focused only on the first two online sessions, since the latter half of

the course differs from faculty to faculty. The content for each session consisted of three video lectures of approximately 10 minutes of length, each followed by a quiz of 4 to 5 multiple-choice questions. In order to analyze students' interaction patterns with 6 video lectures, we recorded videos of the computer screens and participants' faces as they were completing online modules in a laboratory setting. The experiments were run on 17.3-inch laptops. The screen captures and face recordings were later combined into one video and were viewed by two raters who shared the task of rating the participants in terms of playback speed, number of pauses, rewinds and fast-forwards, as well as ratio of wakefulness and number of views. Each parameter has been calculated per slide.

2 RESULTS, DISCUSSION, AND CONCLUSION

We opted for a qualitative approach to finding students' specific behaviors. In an attempt to triangulate the log data with qualitative evidence from students' video viewing behavior, we watched the face recordings and screen captures. Following that, based on our observations, we placed the students into four groups of drowsy, focused, skipping, and mixed-behavior learners. Details of the descriptive statistics of each group along with their quiz scores are displayed in Table 1. Moreover, Figure 1 visually summarizes the four groups' mean number of pauses, rewinds, and fast-forwards as well as their wakefulness and number of views ratio per slide. It is worth noting that the sub-figures each include a primary and a secondary y axis, the primary titled number of times for pausing, rewinding, and fast-forwarding and the secondary titled ratio for wakefulness and number of views ratio values. As can be seen in Table 1 and Figure 1, the first group consisted of drowsy learners, characterized by low levels of wakefulness and low number of views. Particularly, Figure 1 shows that they fell asleep more often in session 4 compared to session 6. This means that there exists an association between wakefulness and difficulty level of videos. As expected, these learners gained relatively low average scores in the quizzes. It is thus necessary to take measures to support drowsy learners by providing appropriate feedback to improve their comprehension and to raise their alertness during e-learning. The second group included learners who were for the most part focused on the screen with minimal number of pauses, rewinds, and fast-forwards. This group outperformed others in the quizzes thanks to their higher focus on the content. The third group was characterized by the highest number of rewinds and fast-forwards. Despite being wakeful, these learners had low number of views and a low mean score in session 4. The last group of learners displayed inconsistent patterns of behavior over the two sessions. For instance, three out of five were drowsy in one session but focused in another, or they only watched one or two videos in one session or frequently fast-forwarded whereas they watched all the videos completely in another session. This group gained the lowest average score in session 4. The results above show that we should detect learners' types and

Table 1: Group and Total Video Viewing Behavior Stats.

Group	Session	Speed	Pauses	Rewinding	Fast-forwarding	Wakefulness	No. of Views	Quiz Scores
1.Drowsy (N=4)	4	1.10 (0.29)	0.00 (0.06)	0.05 (0.28)	0.21 (0.70)	0.59 (0.49)	0.59 (0.54)	44.6 (20.5)
	6	1.17 (0.21)	0.00 (0.06)	0.05 (0.32)	0.03 (0.28)	0.67 (0.47)	0.67 (0.48)	60.0 (9.4)
2.Focused (N=8)	4	1.19 (0.30)	0.01 (0.10)	0.01 (0.15)	0.01 (0.07)	0.99 (0.11)	0.99 (0.13)	57.1 (18.7)
	6	1.48 (0.37)	0.02 (0.13)	0.01 (0.13)	0.06 (0.31)	1.00 (0.00)	0.98 (0.18)	79.2 (17.3)
3.Skipping (N=2)	4	1.04 (0.09)	0.03 (0.17)	0.16 (0.61)	1.26 (1.62)	1.00 (0.00)	0.63 (0.53)	46.4 (5.1)
	6	1.54 (0.33)	0.03 (0.18)	0.22 (0.96)	1.58 (2.11)	1.00 (0.00)	0.68 (0.50)	73.3 (9.4)
4.Mixed (N=5)	4	1.06 (0.17)	0.00 (0.05)	0.00 (0.00)	0.05 (0.23)	0.96 (0.19)	0.84 (0.37)	15.9 (32.9)
	6	1.23 (0.39)	0.00 (0.06)	0.00 (0.00)	0.00 (0.06)	0.94 (0.24)	0.93 (0.26)	69.3 (21.4)

approach them using appropriate feedback. We can detect most types of learners using clickstream data. However, it is difficult to find drowsy learners who also have a higher risk of drop-out. We acknowledge that the current approach is difficult to scale up. Nevertheless, it is an indispensable step in shedding light on learner types and devising ways to approach those in need of support. Given that this blended course is obligatory for all undergraduate students, we aim to improve it in the future by automatizing real-time wakefulness estimation and implementing adaptive learning algorithms capable of tailoring content difficulty and adapting to learners' engagement patterns.

In this study, we analyzed the interaction patterns of learners with video lectures in a blended course. We observed three main learner groups, i.e., drowsy, focused, and skipping, and a fourth group including a mixture of differing interaction patterns. This investigation is a preliminary step in discovering learners' video viewing behaviors in order to create an adaptive learning system that can support various learners, particularly those at risk, and to ultimately lower dropout rates.

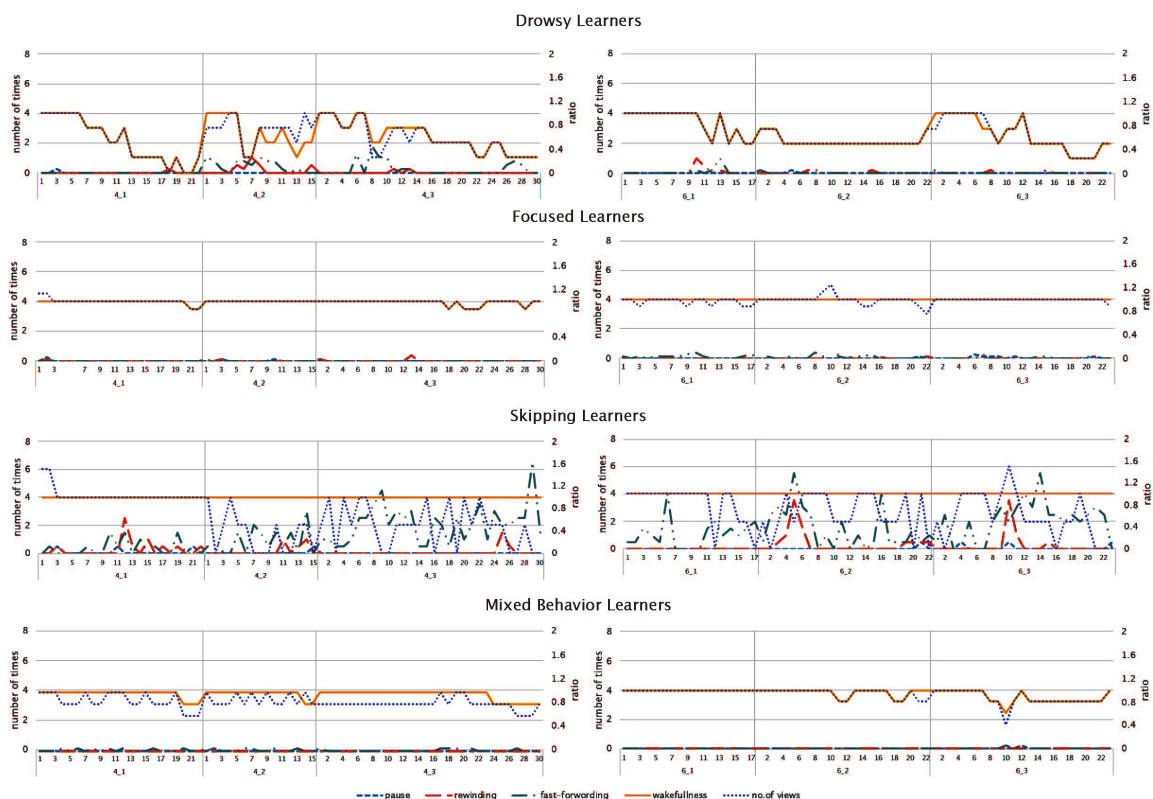


Figure 1: Plots for each group showing the mean metrics per slide

REFERENCES

- Mirriahi, N., & Vigentini, L. (2017). Analytics of learner video use. In C. Lang, G. Siemens, A. Wise, D. & Gašević (Eds.), *Handbook of learning analytics* (pp. 251-267). Society for Learning Analytics Research. <http://doi.org/10.18608/hla17.022>
- Tan, M. & Shao, P. (2015). Prediction of student dropout in e-learning program through the use of machine learning method. *International Journal of Emerging Technologies in Learning*, 10(1), 11-17. <http://doi.org/10.3991/ijet.v10i1.4189>