Companion Proceedings of the 11th International Conference on Learning Analytics & Knowledge LAK20

The impact we make: *The contributions of learning analytics to learning*

April 12-16, 2021

Online,

Everywhere

Organized by:

LAK 2021 Program Chairs' Welcome

We are very pleased to welcome you to the Eleventh International Conference on Learning Analytics and Knowledge (LAK21), organized by the Society for Learning Analytics Research (SoLAR). This year's conference, while originally planned to be hosted by University of California, Irvine at the Newport Beach Marriott, is held virtually April 12-16, 2021 in an effort to protect the LAK community from COVID-19.

The theme for the 11th annual LAK conference is *"The impact we make: The contributions of learning analytics to learning"*. As academic fields concerned with human behavior develop and mature, their impact on advancing scientific understanding and practical application becomes an important marker of success. As an integrated, interdisciplinary and multidisciplinary research field, learning analytics is presented with questions regarding its contributions in two areas: (i) the respective fields from which it draws, and (ii) its own development as a research domain. The LAK21 conference is intended for both researchers and practitioners, and we have invited them to come and join a proactive dialogue around the impact of learning analytics and its practical adoption. We have further extended our invite to educators, leaders, administrators, and government and industry professionals interested in the field of learning analytics and related disciplines.

We received large numbers of high-quality submissions. The research track received 228 submissions (129 full paper submissions and 99 short paper submissions), which represents a slight decrease of about 13% in total number of submissions compared to last year which was to be expected due to the COVID-19 pandemic. It does, however, represent a 7% increase compared to 2019. 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, 36 posters, 4 demos, 8 doctoral consortium submissions, and 18 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. These Companion Proceedings could not have been done without their generous help and support.

We would like to stress our gratitude for the hard effort during COVID times by researchers and all involved in our community. We recognize that as we moved through the past year of the coronavirus pandemic all students, researchers and staff in this community faced new physical and emotional challenges, particularly with feelings of stress, uncertainty and fear. As such, we would like to thank you all for the effort you devoted that has allowed this conference to happen this year.

Our hope is that the LAK21 participants and the readers of these proceedings will recognize the contributions of the field of learning analytics within the scope of the interdisciplinary fields from which we draw. This could include more traditional contributions such as theoretical, methodological as well as practical and community-based contributions such as increased representation from other disciplines (e.g., neuroscience, AI). As an interdisciplinary community ourselves, monitoring these external contributions is particularly meaningful as broadly advancing scientific understanding and practical application becomes an important marker of success within our field.

Maren Scheffel
Ruhr University Bochum

Nia Dowell
University of California, Irvine

Srecko Joksimovic
University of South Australia

George Siemens
University of Texas, Arlington & University of South Australia

LAK 2021 Conference Organization

Program Chairs
Maren Scheffel, Ruhr University Bochum, GER
Nia Dowell, University of California, Irvine, USA
Srecko Joksimovic, University of South Australia, AUS
George Siemens, University of Texas, Arlington, USA & University of South Australia, AUS

Practitioner Track Chairs
Mike Sharkey, Arizona State University, USA
Liz Gehr, Boeing, USA

Workshop and Tutorial Chairs
Caitlin Mills, University of New Hampshire, USA
Paul Prinsloo, University of South Africa, SA
Angela Stewart, Carnegie Mellon University, USA

Doctoral Consortium Chairs
Michael Brown, Iowa State University, USA
Simon Buckingham-Shum, University of Technology Sydney, AUS
Oleksandra Poquet, University of South Australia, AUS
Stephanie Teasley, University of Michigan, USA

Posters and Demonstrations Chairs
Yi-Shan Tsai, Monash University, AUS
Justin Dellinger, University of Texas at Arlington, USA

Social Media Chairs
Elle Wang, Arizona State University, USA
Bodong Chen, University of Minnesota, USA

Website Manager
Yancy Vance Paredes, Arizona State University, USA

Companion Proceedings Editor
Abhinava Barthakur, University of South Australia, AUS

Organizing Committee
Mark Warschauer, University of California, Irvine, USA
Grace Lynch, Society for Learning Analytics Research, AUS
Nicole Hoover, Society for Learning Analytics Research, USA

LAK21 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, DIPF | Leibniz Institute for Research and Information in Education, Germany
Rebecca Ferguson, The Open University, UK
Dragan Gasevic, Monash University, Australia
Neil Heffernan, Worcester Polytechnic Institute, USA
Ulrich Hoppe, University Duisburg-Essen, Germany
Jelena Jovanovic, University of Belgrade, Serbia
Simon Knight, University of Technology, Sydney, Australia
Vitomir Kovanovic, UniSA, Australia
Manolis Mavrikis, UCL, UK
Xavier Ochoa, NYU, USA
Abelardo Pardo, University of South Australia, Australia
Mykola Pechenizkiy, Eindhoven University of Technology, Netherlands
Oleksandra Poquet, University of South Australia, Australia
Bart Rienties, The Open University, UK
John Stamper, Carnegie Mellon University, USA
Stephanie Teasley, School of Information, University of Michigan, USA
Phil Winne, Simon Fraser University, Canada
Alyssa Wise, NYU, USA

Program Committee
Cecilia Aguerrebere, Duke University, USA
Stephen Aguilar, University of Southern California, USA
Nora Ayu Ahmad Uzir, The University of Edinburgh, UK
June Ahn, University of California, Irvine, USA
Carlos Alario-Hoyos, Universidad Carlos III de Madrid, Spain
Vincent Aleven, Carnegie Mellon University, USA
Naif Aljohani, University of Southampton, UK
Kimberly Arnold, University of Wisconsin-Madison, USA
Thushari Atapattu, The University of Adelaide, Australia
Aneesha Bakharia, The University of Queensland, Australia
Tiffany Barnes, North Carolina State University, USA
Alan Berg, University of Amsterdam, The Netherlands
Matthew Bernacki, The University of North Carolina at Chapel Hill, USA
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
Julien Broisin, Université Toulouse 3 Paul Sabatier – IRIT, France
Jill Burstein, Educational Testing Service, USA
Cristian Cechinel, Federal University of Santa Catarina, Brazil
Mohamed Amine Chatti, University of Duisburg-Essen, Germany
Bodong Chen, University of Minnesota, USA
Guanliang Chen, Monash University, Australia
Heeryung Choi, University of Michigan, USA
Irene-Angelica Chounta, University of Tartu, Estonia
Cassandra Colvin, Curtin University, Australia
Miguel Ángel Conde, University of León, Spain

Adam Cooper, Tribal, UK
Linda Corrin, Swinburne University of Technology, Australia
Alexandra Cristea, Durham University, UK
Mutlu Cukurova, University College London, UK
Maarten De Laat, UoW, Australia
Tinne De Laet, Katholieke Universiteit Leuven, Belgium
Carrie Demmans Epp, University of Alberta, Canada
Daniele Di Mitri, DIPF | Leibniz Institute for Research and Information in Education, Germany
Stefan Dietze, GESIS - Leibniz Institute for the Social Sciences, Germany
Yannis Dimitriadis, University of Valladolid, Spain
Shayan Doroudi, Carnegie Mellon University, USA
Brendan Eagan, University of Wisconsin-Madison, USA
Michael Eagle, George Mason University, USA
Vanessa Echeverria, Escuela Superior Politécnica del Litoral, ESPOL University, Ecuador
Erkan Er, Universidad de Valladolid, Spain
Yizhou Fan, The University of Edinburgh, UK
Stephen Fancsali, Carnegie Learning, Inc., USA
Elaine Farrow, The University of Edinburgh, UK
Rafael Ferreira Mello, Federal Rural University of Pernambuco, Brazil
Christian Fischer, University of Tübingen, Germany
Barry Fishman, University of Michigan, USA
Ed Foster, Nottingham Trent University, UK
Stephanie Frost, University of Saskatchewan, Canada
Dilrukshi Gamage, University of Moratuwa, Sri Lanka
Joshua Gardner, University of Michigan, USA
Michail Giannakos, Norwegian University of Science and Technology, Norway
Andrew Gibson, Queensland University of Technology, Australia
Bas Giesbers, Rotterdam School of Management, Erasmus University Rotterdam, The Netherlands
Art Graesser, University of Memphis, USA
Sabine Graf, Athabasca University, Canada
Wolfgang Greller, Vienna University of Education, Austria
David Griffiths, UNIR-iTED, Spain
Xiaoqing Gu, East China Normal University, China
Julio Guerra, University of Pittsburgh, USA
Danielle Hagood, University of California, Davis, USA
Ella Haig, School of Computing, University of Portsmouth, UK
Marek Hatala, Simon Fraser University, Canada
Claudia Hauff, Delft University of Technology, The Netherlands
Christothea Herodotou, The Open University, UK
Isabel Hilliger, PUC, Chile
Martin Hlosta, The Open University, UK
Caitlin Holman, University of Michigan, USA
Helen Hong, Nanyang Technological University, National Institute of Education, Singapore
Sarah Howard, University of Wollongong, Australia
Sharon Hsiao, Arizona State University, USA
Xiao Hu, The University of Hong Kong, Hong Kong
Francisco Iniesto, The Open University, UK
Ioana Jivet, TU Delft, The Netherlands
Rogers Kaliisa, University of Oslo, Norway
Tanja Käser, EPFL, Switzerland
Hassan Khosravi, The University of Queensland, Australia
Kirsty Kitto, University of Technology, Sydney, Australia
Rene Kizilcec, Cornell University, USA
Justian Knobbout, HU University of Applied Sciences, The Netherlands
Elizabeth Koh, National Institute of Education, Nanyang Technological University, Singapore, Singapore

Charles Lang, Columbia University, USA
Eitel Lauria, Marist College, USA
Elise Lavoué, Université Jean Moulin Lyon 3, LIRIS, France
Leon Lei, The University of Hong Kong, Hong Kong
James Lester, North Carolina State University, USA
Tobias Ley, Tallinn University, Estonia
Lisa Lim, UniSA, Australia
Jionghao Lin, Monash University, Australia
Yiwen Lin, University of Michigan Ann Arbor, 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, UK
Vanda Luengo, Sorbonne Université - LIP6, France
Leah Macfadyen, The University of British Columbia, Canada
Bannert Maria, Technical University of Munich, Germany
Mirko Marras, École Polytechnique Fédérale de Lausanne - EPFL, Switzerland
Rebecca Marrone, University of South Australia, Australia
Roberto Martinez-Maldonado, Monash University, Australia
Alejandra Martínez-Monés, Universidad de Valladolid, Spain, Spain
Wannisa Matcha, The University of Edinburgh, UK
Claudia Mazziotti, Technical University of Munich, Germany
Agathe Merceron, Beuth University of Applied Sciences Berlin, Germany
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
Benjamin Motz, Indiana University Bloomington, USA
Kousuke Mouri, Tokyo University of Agriculture and Technology, Japan
Pedro J. Muñoz-Merino, Universidad Carlos III de Madrid, Spain
Hanni Muukkonen, University of Oulu, Finland
Alison Myers, The University of British Columbia, Canada
Ha Nguyen, University of California-Irvine, USA
Quan Nguyen, University of Michigan, USA
Katja Niemann, XING SE, Germany
Rachel Niemer, University of Michigan, USA
Richard Osakwe, Monash University, Australia
Zacharoula Papamitsiou, Norwegian University of Science and Technology, Norway
Luc Paquette, University of Illinois at Urbana-Champaign, USA
Zach Pardos, University of California, Berkeley, USA
Yeonjeong Park, Honam University, South Korea
Melanie Peffer, University of Colorado Boulder, USA
Mar Perez-Sanagustin, Université Paul Sabatier Toulouse III, France
Héctor Javier Pijeira-Díaz, University of Oulu, Finland
Luis P. Prieto, Tallinn University, Estonia
Paul Prinsloo, University of South Africa, South Africa
Mladen Rakovic, Monash University, Australia
Justin Reich, Massachusetts Institute of Technology, 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
José A. Ruipérez Valiente, University of Murcia, Spain
John Saint, The University of Edinburgh, UK
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
Rachel Scherer, Blackboard, Inc, USA
Marcel Schmitz, Zuyd Hogeschool, The Netherlands
Jan Schneider, DIPF | Leibniz Institute for Research and Information in Education, Germany
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
Arabella Jane Sinclair, The University of Amsterdam, The Netherlands
Erica Snow, Imbellus, USA
Sergey Sosnovsky, Utrecht University, The Netherlands
Marcus Specht, TU Delft, The Netherlands
Daniel Spikol, University of Copenhagen, Denmark
Namrata Srivastava, The University of Melbourne, Australia
Angela Stewart, Carnegie Mellon University, USA
Tamara Sumner, University of Colorado Boulder, USA
Zachari Swiecki, Monash University, USA
Yuta Taniguchi, Kyushu University, Japan
Tamara Tate, University of California, Irvine, USA
Michelle Taub, University of Central Florida, USA
Dirk Tempelaar, Maastricht University, The Netherlands
Craig Thompson, The University of British Columbia, Canada
Kate Thompson, Queensland University of Technology, Australia
Stefan Trausan-Matu, University Politehnica of Bucharest, Romania
Yi-Shan Tsai, Monash University, Australia
Anouschka van Leeuwen, Utrecht University, The Netherlands
Ysabella Van Sebille, The University of South Australia, Australia
Lorenzo Vigentini, The University of New South Wales, Australia
Elle Yuan Wang, Arizona State University, USA
Mark Warschauer, University of California, Irvine, USA
Professor Denise Whitelock, The Open University, UK
Alexander Whitelock-Wainwright, Monash University, Australia
John Whitmer, Blackboard, Inc., USA
Steven Williams, University of California, Berkeley, USA
Annika Wolff, Lappeenranta University of Technology, Finland
Marcelo Worsley, Northwestern University, USA
Masanori Yamada, Kyushu University, Japan
Renzhe Yu, University of California, Irvine, USA
Zdenek Zdrahal, The Open University, UK
Chen Zhan, University of South Australia, Australia
Mengxiao Zhu, Educational Testing Service, USA
Amal Zouaq, Ecole Polytechnique de Montréal, Canada

Additional Reviewers
Aaron Alphonsus
Muhammad Anwar
Jason Bernard
Muhammad Chaudhry
Cheng-Yu Chung
Alison Clark-Wilson
Veronica Cucuiat
John Erickson
Mohamed Ez-Zaouia
Gloria Milena Fernandez-Nieto
Michel Gagnon
Ashish Gurung
Aaron Haim
Stuart Halifax
Madiha Khan
Daniel Leiker
Bibeg Limbu
Mathieu Loiseau
Mirko Marras
Mohammed Nehal Hasnine
Korinn Ostrow
Sambit Praharaj
Ethan Prihar
Tamra Ross
Yancy Vance Paredes

Table of Contents

Learning Analytics in the Context of COVID-19: A Case Study of Using Network Analysis to Guide Campus Course Offering Plans

Gina Deom[1], Stefano Fiorini[1], Mark McConahay[2], Linda Shepard[1], Julie Teague[1]
Indiana University-Bloomington, Bloomington Assessment and Research[1]
Indiana University-Bloomington, Office of the Registrar[2]
vpuebar@indiana.edu, mcconaha@indiana.edu

ABSTRACT: Network analysis simulations were used to guide decision-makers while configuring instructional spaces on our campus during COVID-19. Course enrollment data were utilized to estimate metrics of student-to-student contact under various instruction mode scenarios. Campus administrators developed recommendations based on these metrics; examples of learning analytics implementation are provided.

Keywords: COVID-19, Course Enrollments, Network Analysis

1 INTRODUCTION

In response to COVID-19, many universities cancelled Spring 2020 in-person classes and shifted to 100% online instruction. By late spring, institutions began considering plans for Fall 2020. At Indiana University-Bloomington (IUB), planning decisions involving any in-person component were daunting from the start. In a typical semester, over 43,000 students enroll at IUB with more than 2,200 tenured and other faculty teaching and conducting research. Operationally, this yields over 200,000 enrollments across 12,700 classes that take place in roughly 2,300 classrooms. With State guidelines and physical distancing protocols continuously changing, campus committees were tasked with developing recommendations for delivering a safe and effective education. One committee, the Strategic Space Utilization Committee (the committee), was comprised of data analysts, faculty, administrators, registrars, deans, and chairs who focused on ensuring safe use of classroom space (e.g., density reduction) and configuring instruction mode in ways to reduce the number of pathways in which the virus might spread through the student body.

2 DATA-DRIVEN APPROACH

One aspect of the committee's work included a data-driven approach which simulated the interconnectivity of students through performing a network analysis on course enrollment data. This approach was inspired by a Cornell University Study (Weeden and Cornwall 2020) which showed that nearly all students shared a common classmate – a finding that was alarming for colleges trying to reopen their campuses in the COVID-19 environment (Gluckman 2020). Given the implications of the Cornell study, we asked ourselves how we could leverage and expand this work to support the committee. Three network analysis outputs illustrate how this work was applied and used for reopening planning on our campus: 1) performing ad hoc network simulations for initial committee evaluation, 2) exporting class lists to identify potential hot spots in the enrollment network, and 3) constructing a post hoc dashboard to show the impact of decisions on reducing student contact.

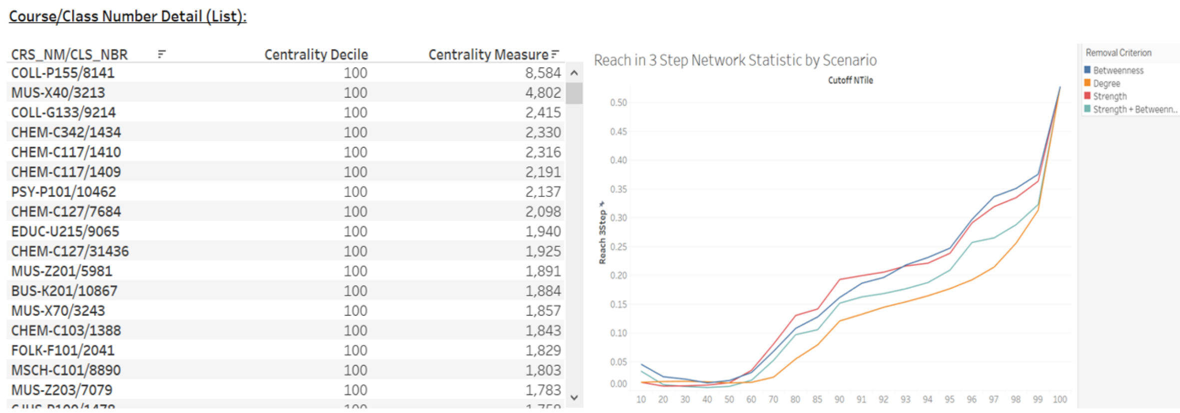
3 RESULTS & IMPLEMENTATION

All three forms of the analysis were highly informative. First, following the approach of Weeden and Cornwall (2020), we discovered that opening campus under pre-pandemic norms would pose significant risk to the university. Using Fall 2019 course enrollment data for fully in-person classes and any hybrid classes that had an in-person component, we found that a typical in-person semester scenario produced an average path length between any two students of 2.7 with a typical student being able to reach 45% of other students within two steps and 79% of students within three steps (see scenario 1 in Table 1). In particular, large lectures appeared to be high contributors of interconnectivity. The committee considered several modality changes related to class characteristics (class size, class component (lecture/non-lecture), etc.) in order to reduce student contact points, and each proposed change was simulated (results for scenarios 2-6 in Table 1). These statistics, especially network reach statistics as they were the most straightforward measures of student-to-student contact, were essential for Fall 2020 planning. As a result of these network analysis scenarios, committee stakeholders recommended a hybrid instruction model with the principle to move certain types of classes online (computer labs, large classes of more than 49 students, multicomponent lectures, and single component general education courses) as defined in scenario 6 in Table 1.

Table 1: Ad Hoc Network Simulation Results Using Fall 2019 Course Enrollment Data

Metric/Scenario	1-All Courses/Sections In-Person	2-Computer Labs Online	3-ClassSize >49 Online	4-Multicomponent Lectures Online	5-Non-multicomponent General Education Online	6-Scenarios 2-5 Combined
# of Edges	7,458,408	7,367,098	1,483,914	4,868,250	5,407,434	1,119,036
# of Vertices	38,386	38,362	37,512	38,374	38,009	35,919
Density	1.0%	1.0%	0.2%	0.7%	0.7%	0.2%
Transitivity	47.6%	48.4%	47.6%	45.8%	56.6%	58.3%
Avg. Path Length	2.74	2.75	3.35	2.88	2.94	3.57
Diameter	16	16	16	16	16	16
2 Step Reach	45.1%	44.6%	9.1%	34.9%	29.9%	5.4%
3 Step Reach	78.7%	78.6%	65.8%	77.0%	76.4%	52.4%

Even though the network simulations showed that such modality changes could reduce spread of the virus, the campus recognized there could still be potential hot spots in the enrollment network. As a second phase to the analysis, we produced network statistics at the class level to identify which classes were the most central in producing connections between students. We also conducted simulations on network reach statistics by removing remaining in-person classes by centrality percentile; this created suggested cutoff points for the class list in terms of moving specific classes to online or hybrid delivery models. Examples of such outputs are provided in Figure 1. This knowledge was particularly informative and actionable for the Registrar's office who was working with departments to set the class schedule and to safely reconfigure remaining in-person classes to available classroom space. Final adoption of the committee recommendations for Fall 2020, largely implemented by the Registrar's office through this ongoing dialogue with academic departments, resulted in a configuration of 63% of the classes offered online and 37% offered in-person or hybrid (not including independent study and research classes). This compared to Fall 2019 in which 93% of the classes were offered in-person or hybrid and 7% were offered online.

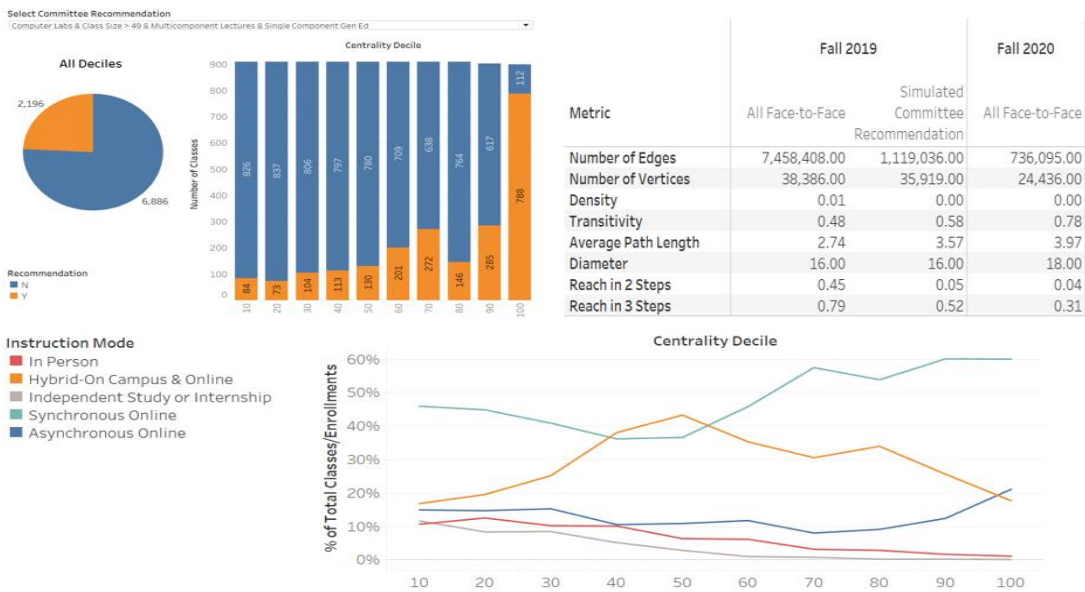


Course/Class Number Detail (List):

CRS_NM/CLS_NBR	Centrality Decile	Centrality Measure
COLL-P155/8141	100	8,584
MUS-X40/3213	100	4,802
COLL-G133/9214	100	2,415
CHEM-C342/1434	100	2,330
CHEM-C117/1410	100	2,316
CHEM-C117/1409	100	2,191
PSY-P101/10462	100	2,137
CHEM-C127/7684	100	2,098
EDUC-U215/9065	100	1,940
CHEM-C127/31436	100	1,925
MUS-Z201/5981	100	1,891
BUS-K201/10867	100	1,884
MUS-X70/3243	100	1,857
CHEM-C103/1388	100	1,843
FOLK-F101/2041	100	1,829
MSCH-C101/8890	100	1,803
MUS-Z203/7079	100	1,783

Figure 1: Excerpts of Network Centrality Class Lists and Reach Statistics by Removal Criterion

Figure 2 shows excerpts of the final stage of the analysis – a dashboard report that highlights the effectiveness of the committee's work: the committee's recommended criteria for online and hybrid instruction picked up a significant proportion of the most central classes (see top left of Figure 2, orange bars); the Fall 2020 student network had significantly fewer student-to-student contact points compared to prior semesters (see top right of Figure 2), and the most central classes in Fall 2020 were also the least likely to have an in-person class component (see bottom of Figure 2). This dashboard report was highly effective in telling the story of the committee's work and quantifying the impact of course modality decisions on reducing student-to-student contact in classrooms.



Metric	Fall 2019 All Face-to-Face	Fall 2019 Simulated Committee Recommendation	Fall 2020 All Face-to-Face
Number of Edges	7,458,408.00	1,119,036.00	736,095.00
Number of Vertices	38,386.00	35,919.00	24,436.00
Density	0.01	0.00	0.00
Transitivity	0.48	0.58	0.78
Average Path Length	2.74	3.57	3.97
Diameter	16.00	16.00	18.00
Reach in 2 Steps	0.45	0.05	0.04
Reach in 3 Steps	0.79	0.52	0.31

Figure 2: Excerpts from Network Analysis Dashboard Report

4 NEXT STEPS

Network analysis provided a structure to facilitate difficult campus-wide discussions and decisions about course delivery. This work was well-received at our institution, and we plan to continue to use this as framework for COVID-19 planning in future semesters. Our campus is exploring ways that we can incorporate other data sources into the network analysis such as on-campus and Greek-life housing data - data that would provide a more complete picture of student interactions. Network analysis could also support the controlled re-introduction of extracurricular activities that are currently online; these activities are an important dimension of a residential campus experience. Finally, we intend to expand the network analysis approach for other purposes; for example, using it to find and explore the benefits of naturally-forming communities at our institution, building on the work of Israel, Koester, and McKay (2020).

REFERENCES

Gluckman, Nell. (2020). ‘A Very Small World’: How Data on Student Enrollment Could Help Colleges Stop Coronavirus’s Spread. *The Chronicle of Higher Education*, 17 April 2020. https://www.chronicle.com/article/a-very-small-world-how-data-on-student-enrollment-could-help-colleges-stop-coronaviruss-spread/

Israel, Uriah, Koester, Benjamin P., & McKay Timothy A. (2020). Campus Connections: Student and Course Networks in Higher Education. *Innovative Higher Education*, 45, 135-151. https://doi.org/10.1007/s10755-019-09497-3

Weeden, Kim A, & Cornwall, Benjamin. (2020). The Small World Network of College Classes: Implications of Epidemic Spread on a University Campus. *Sociological Science*, 7, 222-241. https://sociologicalscience.com/download/vol-7/may/SocSci_v7_222to241.pdf

Diffusion and co-creation: Getting faculty on board a learning analytics platform to personalize student engagement

Danny Y.T. Liu, Natasha Arthars, Nabeel M. Khan, Christopher Pidd, Judy Kay The University of Sydney
Corresponding author: danny.liu@sydney.edu.au

ABSTRACT: Educators typically lack the tools for action-oriented learning analytics to connect with and provide personalized support to students, especially at scale. We discuss educators' adoption of SRES to address these issues. We describe key success factors in implementation and challenges around usability and learnability using a hybrid adoption framework.

Keywords: Adoption; diffusion of innovation; feedback; personalization; student support.

1 BACKGROUND

Student engagement, experience, and learning are strongly influenced by timely academic and personal support as well as frequent, meaningful student-instructor interactions that foster a sense of belonging (Schneider & Preckel, 2017); yet students may feel disengaged and disconnected, due to inadequate support and feedback. Educators understand the stress points in their subjects and are in a prime position to provide relevant monitoring, interaction, and support. However, heavy workloads, scattered and inaccessible data, and a lack of useful technology to act upon that data (West et al., 2015) prevent them from using available data to support students in targeted, educator-driven, and personalized ways, especially at-scale. The Student Relationship Engagement System (SRES) started development in 2012 at The University of Sydney to address this. SRES makes it easier for instructors to capture and make use of relevant, live data from many learning environments (e.g. tutors using SRES's web app to enter grades and feedback while interacting with students; students adding peer- and self-reviews) and curate this with other data (e.g. LMS engagement, performance, participation). It enables educators to control the analyses performed and, importantly, act on data at-scale by creating personalized support for students through 'portals' (programmable web pages that can be made available in the LMS via LTI or standalone) and emails. This content can be static (e.g. written feedback, additional resources, encouragement) and/or interactive (e.g. personalized feedback alongside input fields inviting student reflection). Built-in mechanisms collect engagement metrics and student feedback, helping educators evaluate their actions.

2 IMPLEMENTATION

Since 2012, SRES use has grown at 50% a year, now being used by 1900+ educators to reach 65,000+ students at Sydney University, with growing use at three other Australian institutions. Starting from one faculty, the central learning and teaching unit have supported it from 2016, enabling scaling of support while maintaining its pedagogical and pastoral purposes. The diffusion of innovations (Rogers, 2003) and co-creation (Dollinger, Liu, Arthars, & Lodge, 2019; Payne, Storbacka, & Frow, 2008) models together are informative for analyzing SRES' implementation and adoption; we propose a hybrid framework based on both (Figure 1).

We now explain adoption of SRES in terms of Rogers (2003) who suggested that individuals move through the five stages of the 'innovation-decision' process. Individuals first gain knowledge of the 'why', 'what', and 'how' of an innovation and then decide whether to adopt it based on how they perceive it. This includes its compatibility with their existing practices and values (e.g. supporting students; monitoring participation) and providing relative advantage over existing practices (e.g. digitizing records; scaling personalized feedback) to address felt needs (e.g. reducing workloads, being more connected to students). Being able to trial an innovation on a limited or partial basis assists in a positive adoption decision (Rogers, 2003) (e.g. using SRES for a few weeks just to send personally-addressed messages before exploring other functions). During the implementation phase, Rogers argued that users who modify or change an innovation in that implementation process (e.g. coordinators setting up SRES as a complete student information system) are more likely to continue adopting it sustainably, and that an innovation which lends itself to this so-called 're-invention' is adopted more quickly (Rogers, 2003).

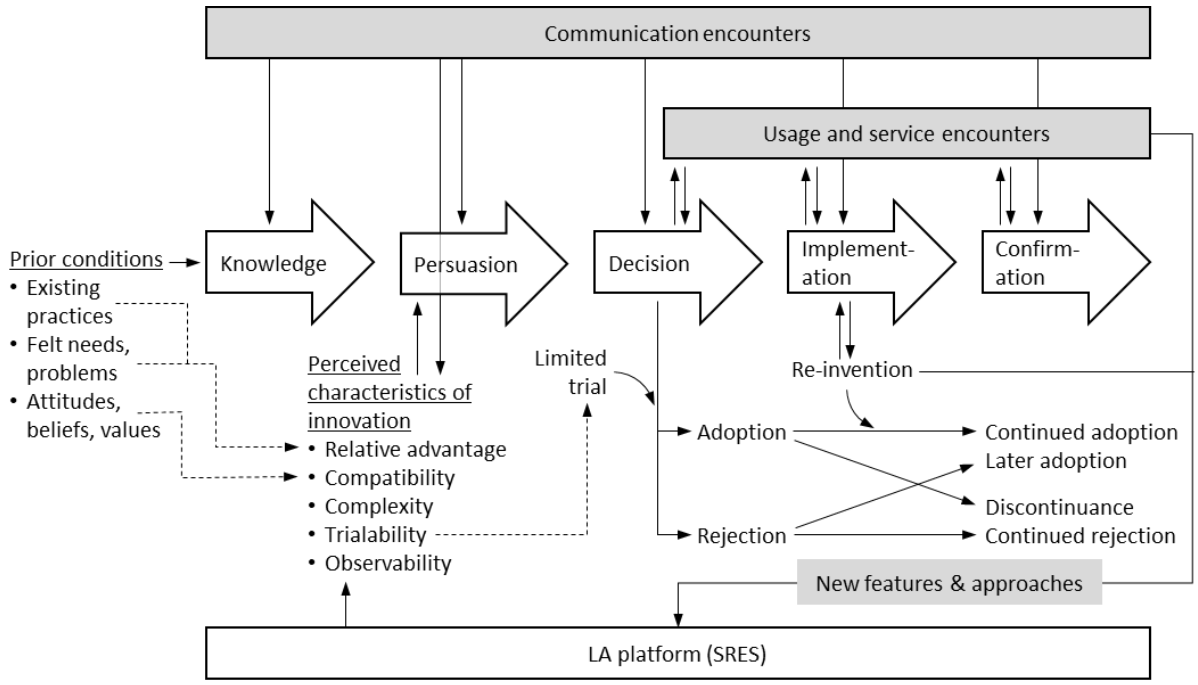


Figure 1: A proposed framework for the adoption of learning analytics platforms by instructors, based on diffusion of innovations (Rogers, 2003) and co-creation (shaded, Payne et al. (2008)) models.

Likewise, the co-creation model by Payne et al. (2008) provides an informative theorization of SRES' adoption by instructors. Their conceptual framework describes how users who are active players in the use, maintenance, and adaptation of a product can co-create 'value', or benefits, for the user (e.g. instructors, in the case of LA, and their students) and the supplier (e.g. universities, learning and teaching units, vendors). Three types of 'encounters' support this process: communication, usage, and service. Communication encounters (e.g. newsletters, announcements) aim to connect with users, in keeping with the importance of communication in Rogers' framework, particularly to support the knowledge and persuasion phases. Interpersonal communication channels (Rogers, 2003) (e.g. peers actively demonstrating benefits and assisting others in overcoming learning hurdles; a 340-member Yammer group) have been key in moving users through the innovation-decision process (Arthars & Liu, 2020). These also support sharing of users' standard and re-invented uses. Usage encounters occur when a product is being used; in the process of implementing a LA platform, users (e.g. instructors, instructional designers) may propose new features (e.g. student data entry, self and peer review, data summaries and dashboards) and once these have been built, users adapted them to creatively expand their uses of the product, going beyond the original design intentions (e.g. providing rich data to teaching assistants to help them better understand students' backgrounds, interests, and progress) (Dollinger et al., 2019). Service encounters occur when users seek training and assistance (e.g. workshops, troubleshooting consultations), and provide another avenue to support the implementation and continued adoption of the innovation.

3 FINDINGS FROM INTERVIEW STUDY

Over eight years, there have been three key challenges for adoption decisions and subsequent implementation: (1) data integration, (2) system stability and responsiveness and (3) system usability, especially learnability. These changed as SRES evolved. For example, data integration issues were largely resolved by Sydney moving to Canvas in 2017; SRES now uses Canvas' open API to access educator-relevant data (e.g. module completions, assessment submissions, site access, discussion engagement). Moving SRES to a modern stack (Python and MongoDB) in early 2019 addressed stability and responsiveness. Usability, including learnability, have improved based on the co-creation processes described above. Some are commonly raised as adoption hurdles, similar to findings by Ali, Asadi, Gašević, Jovanović, and Hatala (2013). Our interviews with 32 SRES users revealed that SRES flexibility led to GUI complexity, cluttering, and navigability challenges. However, users saw significant benefits over existing practices (high 'relative advantage') and high compatibility with existing values and needs. Learnability challenges were noted by some newer users, still in the knowledge or persuasion phases, as they struggled to understand SRES's scope and possibilities; in Rogers' terms, SRES had limited observability. The interviews revealed common software learnability factors such as interface understandability, system guidance appropriateness, and contextual appropriateness (Rafique et al., 2012) (Table 1), and are important for widescale LA software adoption.

Table 1: Examples of how perceived characteristics of an LA platform (SRES) may negatively impact key phases of the innovation-decision process, including representative quotes from user interviews.

	Persuasion phase	**Implementation phase**
High complexity	• Fearful of workload and learning curve, so avoid using the platform • Not confident in being able to understand the platform • "*It was really hard getting people on board because they felt like it was just one more thing to learn and it was easy for [name] because [name] is techy.*"	• Navigability challenges • Documentation length and understandability • "*Figuring out what I wanted to do and at times getting SRES to do what I wanted… I was like 'Why isn't this working?' I was like, 'Oh there's a button I didn't tick that does what I want it to do'.*"

	Persuasion phase	Implementation phase
Low observability	• Difficult to understand how all features of the platform fit with teaching needs • Too many possibilities, where to start? • "*I've had a couple of schools who have looked at introducing it and it's not until they've come in [to] see how we use it that they start to generate their own ideas of how they can use it.*"	• Too many possibilities, unsure which feature to try next • Initial uncertainty of impact • *"(Initially) trying to work out exactly what you want to do and how you want to do it, that's where the difficulty is… [But] if I had to do it without SRES, I would be very busy."*

On the positive side, our interviews indicate SRES' success for the two most important characteristics in innovation adoption (Rogers, 2003), namely relative advantage and compatibility (e.g. "*We've moved from, I'd have to say, not the best feedback mechanisms up to now, to very prompt feedback on any submitted work. So that the students, before they have to complete their next submission task, have an opportunity to improve*"). Indeed, over 1.1 million personalization events (479,421 personalized emails and 648,049 web portal impressions) have been delivered to over 83,000 unique students since mid-2016, with 92.6% (n=6,345 total respondents) of emails and 80.6% (n=14,751 total respondents) of web portals being rated as 'helpful' by students who responded to unobtrusive, embedded feedback prompts. Thematic analysis of student feedback suggests this is due to a heightened sense of connection and care between students and educators, and the provision of timely and tailored feedback. Continued scaling of the use and support of such educator-driven LA tools must balance usefulness (for students and educators) and usability (for educators). This is especially true for flexible systems with diverse functionality and therefore more scope for re-invention. Further research is also needed to understand how educators' use intensifies or degrades over time as they move through the implementation and confirmation phases.

REFERENCES

Ali, L., Asadi, M., Gašević, D., Jovanović, J., & Hatala, M. (2013). Factors influencing beliefs for adoption of a learning analytics tool: An empirical study. *Computers & Education, 62*, 130-148.

Arthars, N., & Liu, D. Y.-T. (2020). How and Why Faculty Adopt Learning Analytics. In D. Ifenthaler & D. Gibson (Eds.), *Adoption of Data Analytics in Higher Education Learning and Teaching* (pp. 201-220). Cham: Springer International Publishing.

Dollinger, M., Liu, D., Arthars, N., & Lodge, J. (2019). Working together in learning analytics towards the co-creation of value. *Journal of Learning Analytics, 6*(2), 10–26-10–26.

Payne, A. F., Storbacka, K., & Frow, P. (2008). Managing the co-creation of value. *Journal of the academy of marketing science, 36*(1), 83-96.

Rafique, I., Weng, J., Wang, Y., Abbasi, M. Q., Lew, P., & Wang, X. (2012). *Evaluating software learnability: A learnability attributes model.* Paper presented at the 2012 International Conference on Systems and Informatics (ICSAI2012).

Rogers, E. M. (2003). *Diffusion of Innovations, 5th Edition*: Free Press.

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological bulletin, 143*(6), 565-600. doi:10.1037/bul0000098

West, D., Huijser, H., Lizzio, A., Toohey, D., Miles, C., Searle, B., & Bronnimann, J. (2015). Learning Analytics: Assisting Universities with Student Retention, Final Report (Part 1). In: Australian Government Office for Learning and Teaching.

Leveraging Courseware Engagement Data to Improve Student Success

Maureen A. Guarcello
San Diego State University
mguarcello@sdsu.edu

James P. Frazee
San Diego State University
jfrazee@sdsu.edu

Asa Levi
Pearson
asa.levi@pearson.com

Jason Lokkesmoe
Pearson
jason.lokkesmoe@pearson.com

Tristan Hillis
San Diego State University
tristan3214@live.com

ABSTRACT: This collaborative study focused on observing San Diego State University students' performance, combined with engagement data generated from Pearson's MyStatLab homework platform within a high enrollment, historically challenging statistics course. Engagement patterns and potential risk were identified using clustering methods, ultimately forecasting students' course completion path before assessments were administered.

Keywords: Predictive Modeling, Learning Analytics, Data Informed Intervention, Student Success

1 BACKGROUND

This study was conducted using data from a statistics course at San Diego State University (SDSU). In response to a strategic planning charge, SDSU convened a learning analytics working group which included faculty, administrators, and staff. Nearly a decade later, the group's shared goal to leverage student performance and engagement data to increase student success in historically challenging courses has continued, now a key component of the four and six-year graduation rate initiatives at the campus and system levels.

A ubiquitous obstacle in this research was accounting for students' engagement with publisher courseware, which faculty use to administer homework, quizzes, and reading assignments. While the learning management system data reflected students' navigation to the content once they logged on to the courseware portal, SDSU was unable to view the activity (e.g., time on task, question attempts, etc.). Likewise, publishers had access to student activity data, but no indication of final course grades.

Without these complementary data points, neither organization had a complete picture of the students' learning path, nor could they leverage the data to increase performance through strategic interventions (e.g., Supplemental Instruction, tutoring, office hours, etc.). In order to overcome the disconnect, SDSU and Pearson partnered in Fall 2018 to investigate whether they could gain insights into student outcomes by leveraging performance data from the content platform itself.

2 DESCRIPTION

Throughout the past five years, repeatable grade rates (C- through D, F, or withdrawals) in SDSU's largest, lower-division introductory statistics courses have been >23%. Performance challenges are compounded by the >1,000 students enrolled in the course each semester. Some trends were identifiable, showing promise towards catching students before they failed. A closer look at students' exam one outcomes over three semesters demonstrated that more than 60% of students who failed the first exam ultimately received a D or F in the course. This early signal of students' success or failure set the foundation for the study.

In Fall 2018, SDSU partnered with Pearson to explore data generated by students on the MyStatLab courseware platform with hopes of improving statistics course performance outcomes. A repeatable method was developed using only behavioral data; a strategy that had two objectives. The goals were to leverage data that students were able to act upon (i.e., no demographic or prior performance data), and to ensure stable algorithmic predictions in the first weeks of the semester, allowing for interventions before the first exams were administered.

The following research questions guided the inquiry:

Can historic, course-level data be used to build a model which identifies student engagement patterns within required course activities (e.g., homework, quizzes, class participation)?

How early in the semester (15 weeks of instruction) might a students' course engagement serve as an indication of their overall course performance?

How can SDSU and Pearson leverage the engagement data and provide strategic interventions to increase students' success?

3 RESULTS

The study began with an exploration of student courseware engagement patterns. A majority of students' coursework in the class was attempted after faculty and peer assistance were no longer available, and in the final hours before the assignments were due. Looking at the data and distinct engagement patterns, a clustering model was used to fit the data. The first attempt was to group data by day, and to cluster using engagement duration, sessions, assessments, and questions. While the method worked, grouping sessions into multi-day buckets ultimately provided the cleanest signal. From that method, the following four canonical behavior patterns were demonstrated by students in the statistics class.

1) *Early Compliant* students initiated and completed their homework assignments three or more days before the deadline, demonstrating a distributed study practice. 2) *Compliant* students typically began their homework 24-48 hours in advance of the deadline. 3) *Late Compliant* students began and submitted work on the due date. While there are more "high risk" students in this cluster, the performance distribution is similar to the Compliant group. 4) *Non-Compliant* student engagement patterns indicated fewer submissions, and those completed were often submitted past the due date.

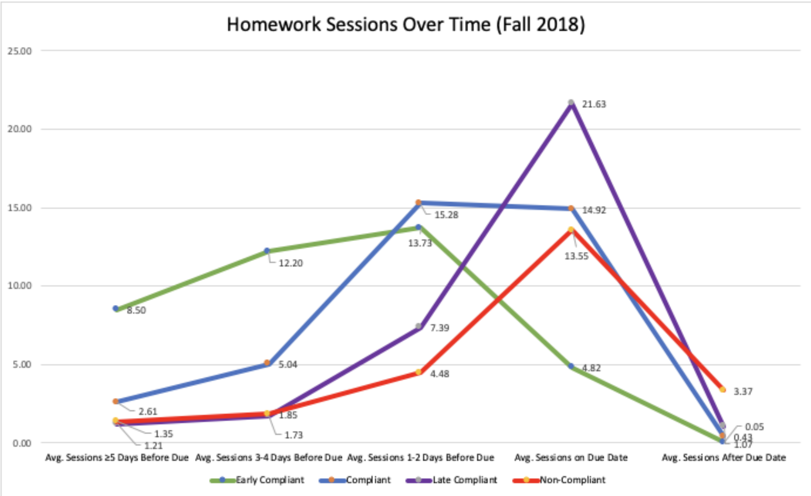


Figure 1. Four clusters depicted from Early Compliant through Non-Compliant, based upon average student interactions with the MyStatLab homework platform.

By clustering students based upon engagement patterns, the predictions proved to be an equitable and accurate metric to identify at-risk students early in the semester. At the end of week two, the accuracy of the model was 52%, with an observed sensitivity of 38%, which was 65% better than a random guess. By week nine, 86% of students were accurately clustered based on their engagement patterns. The first exam is typically administered in week six, providing a month-long head start to connect at-risk students to the resources they need to be successful. These results were encouraging, indicating that it is possible to identify behavior patterns early in the semester with enough accuracy to enable directional communication helping students succeed.

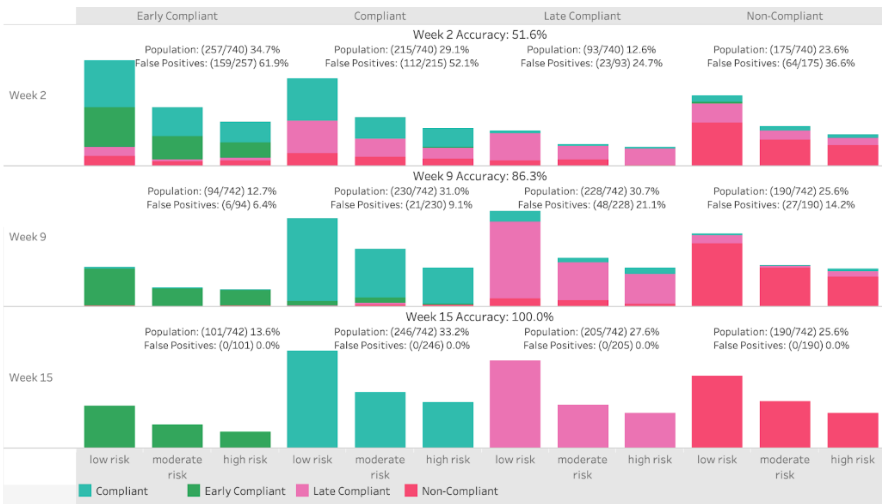


Figure 2: Student observations by Compliance cluster at weeks 2, 9, and 15.

4 FUTURE WORK

Pearson and SDSU hope to partner again to scale this methodology and provide guided interventions from the second week of a course through the first midterm, a critical time period. The idea is to use these data-informed trends to reinforce available student success pathways. SDSU's Supplemental Instruction program has demonstrated success since its introduction in 2015, improving student exam scores by a full letter grade, for those who attend the peer-facilitated study sessions. Targeted communications highlighting the availability of these resources may increase students' chances of passing, especially those students in the Late and Non-Compliant clusters.

The true test of the methodology is the capacity for scale and sustainability across subject areas. Pearson hopes to expand the application of this methodology to other courses and universities. Of particular interest is investigating the viability of the approach in different educational contexts, including continuing education, or in courses with multiple weekly assignments.

REFERENCES

Arendale, D.R., (1994). *Understanding the Supplemental Instruction model*. New Directions for Teaching and Learning, 60, 11-21.

Asilomar Convention for Learning Research in Higher Education. (2014). Retrieved from http://asilomar-highered.info

Guarcello, M.A., Frazee, J.P., Levi, A., Lokkesmoe, J., & Hillis, T. (2020). *Forecasting the journey, not the destination: Leveraging courseware engagement data to improve student success* [White paper]. Retrieved October 15, 2020, from https://its.sdsu.edu/docs/SDSU_Pearson_White_Paper_20.pdf

Smarr B.L., Schirmer, A.E., (2018). *3.4 million real-world learning management system logins reveal the majority of students experience social jet lag correlated with decreased performance.* Scientific Reports, 8, 1-9.

Personalized Learning Analytics-Based Feedback on a Self-Paced Online Course

Joonas Pesonen[1], Outi-Maaria Palo-oja[2] & Kwok Ng[2,3]
[1]University of Helsinki, [2,3]University of Eastern Finland, [3]University of Limerick
joonas.pesonen@helsinki.fi, outi-maaria.palo-oja@uef.fi, kwok.ng@uef.fi

ABSTRACT: Self-regulated learning skills are very important on self-paced online courses, where students have great autonomy but limited guidance. We combined goal setting strategies and personalized feedback to support students' self-regulation and keep them engaged in a self-paced online course. We describe our setup and the findings of the project.

Keywords: self-regulated learning, goal setting, OnTask

1 BACKGROUND

One of the most promising applications of learning analytics is to use data to scale the provision of personalized feedback to students (Pardo et al. 2017). Getting personalized feedback with this approach has been associated with a positive impact on student perception of feedback quality (Pardo et al. 2017) and higher academic achievement (Lim et al. 2019). Moreover, Lim and colleagues (2019) found that patterns of self-regulated learning (SRL) differed between students who received personalized feedback and those who did not.

Promoting students' SRL skills will increase their chances of success in online learning. The importance of SRL is highlighted in self-paced setting. Whereas an imposed-pace model stipulates that all learners engage in the same learning activities at specific time periods, the self-paced approach affords more autonomy to learners (Rhode 2009), increasing the need for self-regulation.

When applying personalized feedback in a self-paced setting, the feedback process must be designed in a manner that takes each student's pace into account. This can be achieved by asking students to set their goals at the beginning of the course. Goal setting is one of the SRL subprocesses and higher application of goal setting has been associated with increased course completion (Handoko et al. 2019). The goal setting activity can both act as an SRL intervention and provide data for personalization: students can be given feedback in respect to the goals they have set.

2 IMPLEMENTATION

We combined goal setting and personalized feedback on an undergraduate level online business course that students could enrol and complete throughout the academic year. The course included three modules which were each assessed with an online exam (30 or 35 points each, total 100 points). In previous years, students often registered on the course early on but actually started studying only when the end of semester was approaching. The aim for the academic year 2020-2021 was to reduce the average course completion time as well as increase the completion rate.

The pilot took place in the fall semester 2020. After enrolment, students set their goals by stating which month they intend have the course completed and consented to receive personalized feedback. We used *OnTask* (Pardo et al. 2017) to provide students with personalized feedback. Feedback rounds were carried out twice a month and, on each round, messages were created for meaningful combinations of completion goals (5 alternatives) and course progress (6 alternatives). During the course, each student would get on average three messages encouraging the student to keep on track with the completion goal. An example of *OnTask* is in Figure 1.

> Hi {{ firstname }}!
>
> {% if november %}{% if welcome questionnaire completed, first exam not completed %}
> You planned to complete [redacted] course in November. Usually it takes about three weeks to complete one module If you start now you may complete the course as planned. You will find slides and test quiz for the first module here. The next two modules will open as the previous one is passed. Good luck!

Figure 1: Using completion goals and course progress as conditions in OnTask.

3 FINDINGS

By the end of October 2020, 258 students had enrolled into the course (156 on previous year). During the first two months more students completed the course (N=21; 8 % of enrolled students) compared to the previous year (N=2; 1 % of enrolled students).

Majority of students (77% - 81%, depending on the feedback round) were willing to receive personalized feedback. Based on the learning management system logs, online activity increased by two to three times the usual activity during the days when the feedback was sent. Furthermore, 26 students replied to the feedback messages by reflecting on their completion goal, for example, stick to the plan, delay completion, or withdraw from the course.

These findings give an early indication that goal setting and personalized feedback supported students' self-regulatory behaviour on the course. Although we cannot be sure if it is due to this intervention, the completion rate during the first two months was higher than on the previous year, which looks promising to meeting the project aims.

REFERENCES

Handoko, E., Gronseth, S. L., McNeil, S. G., Bonk, C. J., & Robin, B. R. (2019). Goal setting and mooc completion: A study on the role of self-regulated learning in student performance in massive open online courses. *International Review of Research in Open and Distributed Learning, 20*(3).

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*, 101202.

Pardo, A., Jovanovic, J., Dawson, S., Gašević, D., & Mirriahi, N. (2017). Using learning analytics to scale the provision of personalised feedback. *British Journal of Educational Technology, 50*(1), 128-138.

Rhode, J. (2009). Interaction equivalency in self-paced online learning environments: An exploration of learner preferences. *The international review of research in open and distributed learning, 10*(1).

Supporting Study Behavior with the THERMOS Learning Analytics Dashboard

Lars de Vreugd
Centre for Research and Development of Education, University Medical Centre Utrecht, The Netherlands
L.b.devreugd-2@umcutrecht.nl

Renée Jansen
Centre for Research and Development of Education, University Medical Centre Utrecht, The Netherlands
R.s.jansen-14@umcutrecht.nl

Anouschka van Leeuwen
Department of Educational Sciences, Utrecht University, The Netherlands
A.vanleeuwen@uu.nl

Sergey Sosnovsky
Department of Information Sciences, Utrecht University, The Netherlands
S.a.sosnovsky@uu.nl

Marieke van der Schaaf
Centre for Research and Development of Education, University Medical Centre Utrecht, The Netherlands
M.f.vanderschaaf-4@umcutrecht.nl

Abstract: The THERMOS dashboard provides students with actionable feedback regarding the following aspects of study behavior: motivation, engagement, groupwork skills and progress. These aspects are visualized on a learning analytics dashboard, along with suggestions for increasing these aspects. Initial results show users' interest in the dashboard and their need for explanations of the dashboard's context and purpose. Detailed data pertaining to usability and perceived value is currently being gathered to be presented during the conference.

Keywords: Learning Analytics Dashboard, Motivation, Study behavior, Student support

PROJECT BACKGROUND

Student dropout rates in higher education in the Netherlands are substantial (Vossensteyn et al., 2015). Dropout rates and study progress are influenced by study behavior and motivation. Improvement of study behavior and motivation may be achieved by feedback on these constructs (Hattie & Timperly, 2007), which can be provided to students through a learning analytics dashboard. THERMOS is such a dashboard (Figure 1). On the dashboard, students first fill in a questionnaire to self-assess their motivation, engagement (Motivation and Engagement Scale, MES; Martin, 2016) and group work skills (Groupwork Skills Questionnaire, GSQ; Cumming et al., 2015).

This data is visualized to students in separate graphs (Figure 1, parts 1 and 2), along with real time study progress data (i.e. credits obtained and GPA) (Figure 1, part 3). A feedback box (Figure 1, part 4) provides students with actionable feedback, consisting of follow-up activities within the dashboard (e.g. exercises to improve planning) and additional resources (e.g. getting in touch with a

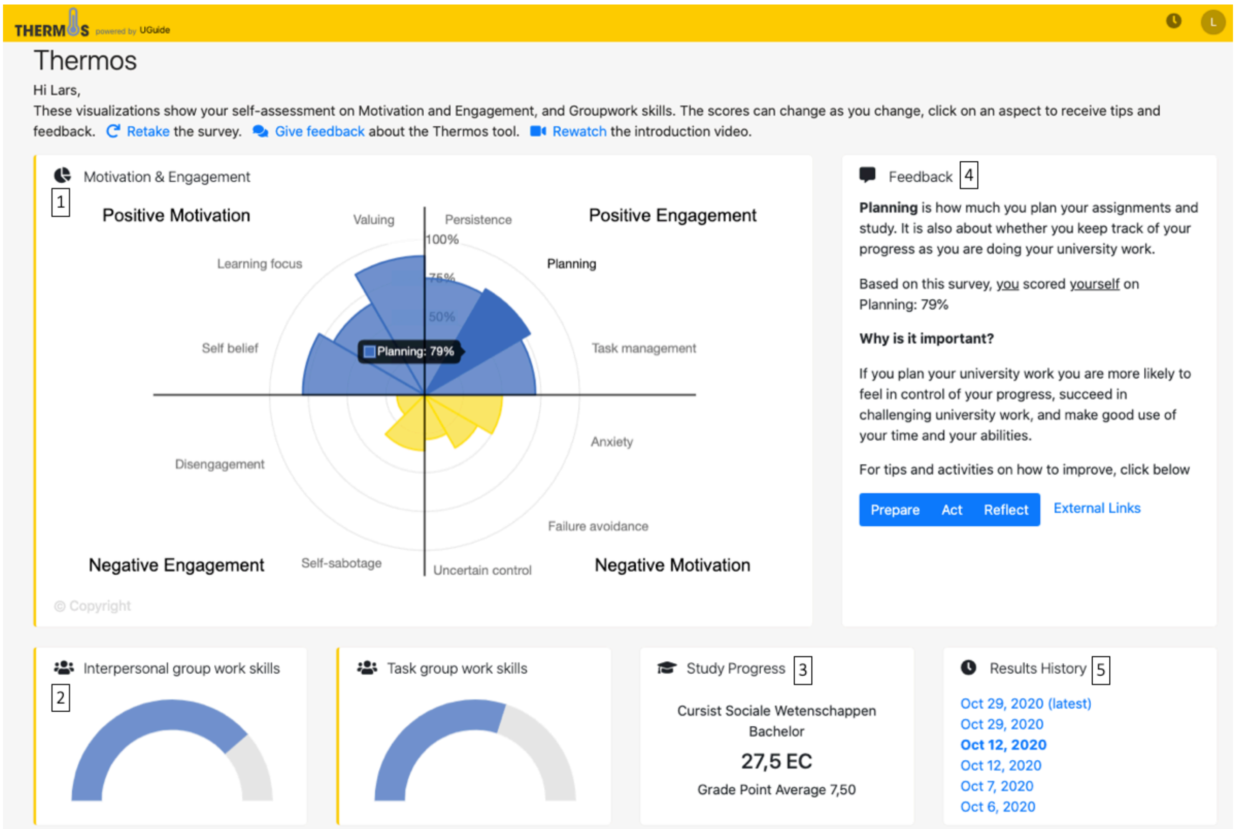


Figure 1. Dashboard screenshot with 1) MES visualization, 2) GSQ visualization, 3) Study progress, 4) Feedback box, and 5) Results history.

tutor). Actionable feedback is based on available materials (e.g. MES Workbook, Martin, 2016), available university workshops, and collaboration with other university projects. The results' history widget (Figure 1, part 5) allows students to access their past data entries, allowing for comparison or reflection on personal growth.

IMPLEMENTATION DESCRIPTION

The dashboard is intended for all faculties and is implemented in several tutor programs at Utrecht University, The Netherlands. Both its development and evaluation took place with several groups of stakeholders (i.e., students and tutors) to gain valuable bottom-up information.

To inform the dashboard's design and align its content with users preferences, separate focus groups with students (total *n*=16) and tutors, teachers, and study advisors (total *n*=7) were organized at the start of the project (academic year 18/19). Here, participants were informed about the dashboard's goal and asked what constructs they deemed important to incorporate. The nominal group technique was used, which allows for an individual thought process before group comparison. A list of possible constructs, based on literature, was also presented for participants to reflect upon. During follow-up sessions, participants reflected on dashboard design and drew mockups. Several examples (e.g. spider graph) were shown to inform them of technical possibilities. After a literature search for existing frameworks and measurement methods, aforementioned aspects (e.g. motivation) were selected. The initial design was based on participant input, as well as insights from literature (e.g. Van der Schaaf et al., 2017).

The evaluation of the dashboard began at the end of academic year 19/20. A small-scale usability test was performed (*n*=8 students), before implementing the dashboard on a larger scale. This study aimed to obtain an indication of students' perceptions of usability and usefulness regarding the dashboard. This pertained to student's understanding of the dashboard and its goal, its design (e.g. interpretation of graphs), and specific functions (e.g. hovering over constructs to receive feedback). Students performed tasks in the dashboard using fictitious profiles, e.g. "Please rank [Profile]'s top three skills/attributes, where number 1 is his/her strongest. Please think aloud while doing so". Statements and tasks performed were analyzed and compared to explore patterns and determine students understanding.

These results led to an adjustment of the dashboard, which was further evaluated in academic year 20/21. Here, the dashboard was implemented in diverse study programs, approximately 400 students participated in iterations of one semester (two in total). The first iteration led, similar to the small-scale usability test, to a refinement of the dashboard. Data from the second iteration aimed validate students' perceived usability and usefulness of the dashboard.

FINDINGS FROM PROJECT AND PRODUCT EVALUATION

Throughout the development of the dashboard there was a close collaboration between developers, tutors, and students. The design was informed by end users' insights and feedback, which added to its usefulness in educational practice.

A striking finding from the first small-scale evaluation was that although students understood the dashboard's functionalities (e.g. graphs and buttons), they had difficulty understanding the value of the dashboard even though they were briefed on the goals and function of the dashboard at the start of the usability test. Participants from this small-scale test were also likely more motivated to use the dashboard than the general student population, as they volunteered to participate in a study for which they received no immediate compensation. This finding highlights the importance of clearly explaining the dashboard's function, goals, and (perhaps most important) the possible value to users. An example to address this issue, was the addition of a tutorial video in the first evaluation iteration of 20/21, explaining the dashboard and its possible value. Also, in the communication towards students (e.g. via email), the dashboard's goal and potential value was already briefly indicated. Furthermore, implementing the dashboard in tutor programs for both 20/21 evaluation iterations helped provide context for students. This in turn added to the understanding of the dashboard's value and making it an inherent part of the study program. This way of implementing is labor intensive however (for both researchers and tutors), as tutor programs' structures and contents vary across faculties. The trade-off seems to be worthwhile, as it adds to the integration of the dashboard and students understanding of the dashboard's potential value.

In conclusion, by actively involving end-users (i.e., students) and educational staff (tutors, teachers, and study advisors) the project gained valuable bottom-up information during the design of the dashboard. The first evaluation of students' perceptions of usability and usefulness allowed a refinement of the dashboard before larger implementation in tutor programs took place. Further insight in usability and usefulness came from the first evaluation iteration in 20/21, leading to the final version evaluation in iteration two. In the future, further integration of THERMOS in tutor programs (and possibly allowing access) may also enhance tutors' assessment of students, as it provides another source of information pertaining to students' needs in monitoring their own study behavior and progress. Furthermore, tailoring the dashboard to end users' needs is a future aim as recent learning analytics literature has highlighted the importance of doing so. Taking into account students' motivational dispositions (Schumacher & Ifenthaler, 2018), their goal achievement orientation (Beheshita et al., 2016), or adding customizable components (Wise & Vytasek, 2017) could be worthwhile endeavors. Ideas for tailoring the dashboard (e.g. adding reference norms or decision-making support) and data from the second evaluation iteration will be discussed during the conference.

REFERENCES

Beheshitha, S. S., Hatala, M., Gašević, D., & Joksimović, S. (2016, April). The role of achievement goal orientations when studying effect of learning analytics visualizations. In *Proceedings of the sixth international conference on learning analytics & knowledge* (pp. 54-63).

Cumming, J., Woodcock, C., Cooley, S. J., Holland, M. J., & Burns, V. E. (2015). Development and validation of the groupwork skills questionnaire (GSQ) for higher education. *Assessment & Evaluation in Higher Education*, *40*(7), 988-1001.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81-112.

Martin, A.J. (2016). The Motivation and Engagement Scale (16th Edition). Sydney, Australia: Lifelong Achievement Group (www.lifelongachievement.com).

Schumacher, C., & Ifenthaler, D. (2018). The importance of students' motivational dispositions for designing learning analytics. *Journal of Computing in Higher Education*, *30*(3), 599-619.

Van der Schaaf, M., Donkers, J., Slof, B., Moonen-van Loon, J., van Tartwijk, J., Driessen, E., Baddi, A., Serban, O., & Ten Cate, O. (2017). Improving workplace-based assessment and feedback by an E-portfolio enhanced with learning analytics. *Educational Technology Research and Development*, *65*(2), 359-380.

Vossensteyn, J. J., Kottmann, A., Jongbloed, B. W., Kaiser, F., Cremonini, L., Stensaker, B., ... & Wollscheid, S. (2015). Dropout and completion in higher education in Europe: Main report.

Wise, A. F., & Vytasek, J. (2017). Learning analytics implementation design. *Handbook of learning analytics*, *1*, 151-160.

iFest 2020: Learning Analytics of a Virtual Event

Biljana Presnall
Jefferson Institute
bpresnall@jeffersoninst.org

ABSTRACT: The Federal E-learning Science & Technology (iFest) conference, an important annual event for the military training and simulation community organized by the National Training & Simulation Association (NTSA), was threatened with cancelation this year because of the pandemic. It normally is attended by about 300 individuals, who can earn Continuing Learning Points (CLP) certified by the Defense Acquisition University (DAU). To save this year's conference, two volunteer groups transformed it into a virtual event, which attracted 450 participants. iFest 2020 included a learning analytics strategy utilizing Experience API (xAPI)[1], which created a clear pathway for meeting the continuing education requirements, enhanced overall participant experience in the virtual conference, and delivered actionable insights for exhibitors, program committees, and presenters.

Keywords: virtual conference, continuing education points, xAPI, learning analytics

CHALLENGES OF THE VIRTUAL EVENT

iFest is one of several key annual events for the military training and simulation community, but the pandemic made an in-person conference impossible. Lacking the budget and technical capacity to transform it into an online event, the organizing committee turned to a pro-bono team to create a virtual experience for iFest 2020. The conference took place on August 17-19, 2020. The live sessions included two keynote speakers (one military and one government), the War on the Rocks[2] podcast, and panel discussions. The pre-recorded sessions were organized around eight themes with a total of 50 full presentations and 10 poster sessions. Chatrooms solved the main challenge of the event: enabling live interaction between the participants. Eight chatrooms matched the pre-recorded session themes, one "attendee lounge" hosted general discussions, and there was one chatroom for the live sessions. In addition, a forum was open for each of the presentations, posters, and live sessions. The exhibit hall featured an independent page for each of the ten vendor exhibitors, who delivered presentations that participants could rate. Although participants could access the platform before and after the event, the learning points and participant interaction occurred only on the conference days, which is when the 450 registered users were most active: 447 participated on the first day, 276 on the second, and 184 on the third.

[1] https://adlnet.gov/projects/xapi/

[2] https://warontherocks.com/

ANALYTICS WITH XAPI

We wanted to use xAPI technology to trace participants' granular behavior during the event, but we had a problem: unlike some ready-made services for website analytics, there is no off-the-shelf xAPI solution for a virtual conference. To transform all the richness of an in-person event into a click, we used JavaScript library xAPIWrapper,[3] which manages communication with a Learning Record Store (LRS). We had to repeat insertion of the xAPIWrapper on every page of the conference website, using the coding steps below. (Figure 1)

```
include library wrapper.js
conf = parameters for the lrs
triger_loc = each loc we need to send info to lrs
foreach_triger_loc
call create_statement
function create_statement
  CASE action OF
     "open page": verb = "launched"
     "click link": verb = "attended"
     "post comment": verb = "commented"
       ...
   ENDCASE
   object = learning object
   actor = user email
   context = context
```

Figure 1: Pseudo-code for generating xAPI statements

With the library enclosed in the object, we could use verbs to create a path to record meaningful action the attendees took on the website. We used the following set of verbs: attempted, attended, commented, completed, exited, experienced, initialized, interacted, launched, preferred, played, resumed, shared, suspended, and terminated. Because xAPI does not collect data anonymously, participants acknowledged and consented to interaction upon entering the platform. In the background, the xAPI statements were sent to an external LRS for collection and retrieval.

Within the conference website on a dedicated analytics page called KUDOS, xAPI enabled five visualizations: most viewed on-demand presentation, most comments, most active in chats, highest rated presentations, and the leader board of "addicted to iFest" participants. The KUDOS page was dynamically generated based on participants' interaction, and it displayed top videos and participants, which encouraged frequently checking back on the page during the conference for updates. On average, the participants checked the current update on the KUDOS page after every 4.6 interactions on the platform.

[3] https://github.com/adlnet/xAPIWrapper

We obtained 26,870 xAPI statements during the three conference days. Generating xAPI statements stopped at noon on the event's last day when the winners of Best Poster Design and Best Poster Narrative were announced. This process of selecting the winners was automatic, based on the xAPI data, and the outcome was uncertain until the end, when the same poster won both prizes. This prompted some participants to argue for a non-automatic process in favor of a more "democratic" approach.

The xAPI data also was used to automatically issue Continued Learning Points (CLP) from the Defense Acquisition University (DAU). DAU had approved a maximum of nine CLPs for attendees who met three requirements: participating in all three main event days; viewing a minimum of four pre-recorded sessions; and actively participating in live chat discussions or forums. The platform awarded the certificates for nine CLPs to only 92 participants. We had expected a higher number of attendees to achieve this benchmark, and we discovered that "viewing" the video presentations was the main problem. Only the participants who watched a video all the way to the end were credited with having viewed it. The data shows that considerably more attendees glanced at a video presentation or watched it partially than those who viewed it in its entirety. (Figure 2)

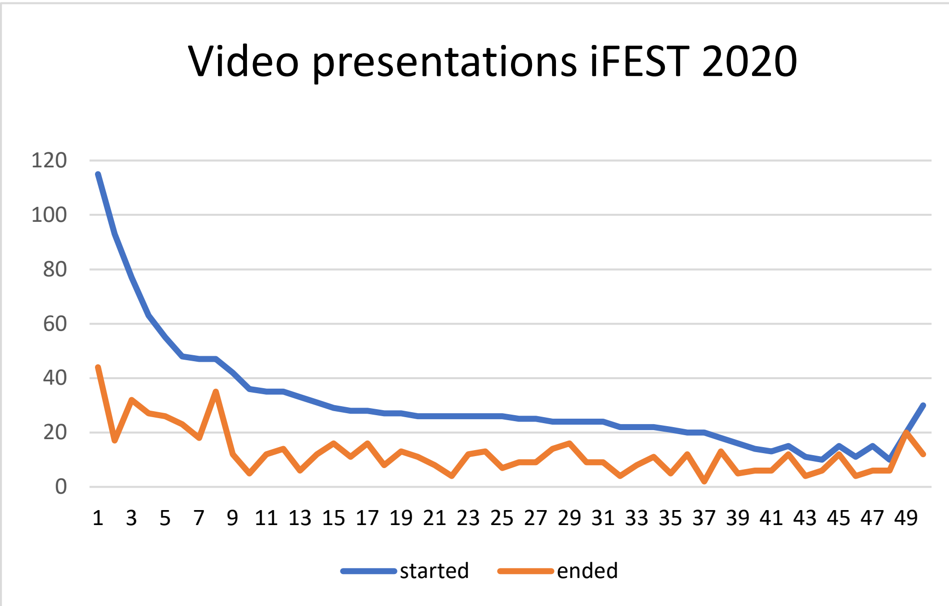


Figure 2: Video presentations: started vs. ended

Given that some of the presentations were longer than 30 minutes, it is not surprising that viewers watched only part of them. However, this diminished the effective utilization of the xAPI video profile's full potential for integrating participant viewing of the pre-recorded presentations as part of CLP requirements. Setting "viewing" on a different level that depends on more granular learning goals would more accurately align with the CLP issuing requirements.

Deeper text analysis of the xAPI data reveals that the participants expressed an overall positive sentiment toward the virtual conference. Despite the difficulties inherent in converting such a large and established event to online-only, and essentially at the ultimate hour, it made for a pleasant conference experience with multiple options for participant interaction. In addition to the chatrooms and discussion forums facilitated by authors and vendor exhibitors, we also had two unexpected interaction opportunities: Book Club, where attendees could recommend a reading; and Matchmaking Board, where participants could share or seek something.

With the lessons we learned from iFest 2020, we are working to develop an xAPI virtual conference profile which would ease the process of organizing and conducting similar events in the future. With the pandemic still preventing full-time, face-to-face gathering, we might have many more opportunities to use it than we imagined a year ago.

KEY LESSONS LEARNED

- Video analytics: The 50 pre-recorded video sessions were a major part of the virtual experience, but the interaction was downgraded due to the setup of the view count. More complete insights about a video's popularity requires relating view count to other preference methods. (e.g., video xAPI profiles, comments, ratings, etc.)

- Heuristic development approach: The mixed military-civilian audience appreciated the early investment into a heuristic evaluation of the virtual event, which ensured ease of use and security supported by a strong learning concept strategy (Jackson 2019).

- When iFest was reconceptualized as a virtual learning event, learning analytics was a natural tool kit to leverage; however, it is important to continue refining the approach to fully demonstrate the LA benefits for both participants and organizers.

REFERENCES

Chatzopoulou, G., Sheng, C., & Faloutsos, M. (2010). A First Step Towards Understanding Popularity in YouTube. Proceedings of the INFOCOM

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. *Journal of Learning Analytics* (1), 120-144.

Jackson, D. (2019). *Design by Concept: A New Way to Think about Software*. Independently Published

Nielsen, J., & Molich, R. (1990, March). Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 249-256.

Enhancing AI-enabled Adaptive Learning System with Sub-skills Modelling

Ioana Ghergulescu, Conor O'Sullivan
Adaptemy
ioana.ghergulescu@adaptemy.com, conor.osullivan@adaptemy.com

ABSTRACT: Adaptemy AI Engine powerfully applies AI on curriculum, content and assessment to enable effective teaching and learning. AlgebraKiT Engine analyses the steps students take when solving maths problems and offers immediate hints to students. This project integrated the two engines to enhance personalised adaptive maths practice for secondary level students.

Keywords: AI-enabled Adaptive Learning System, learner model, sub-skills modelling.

1 BACKGROUND

The need to enable effective teaching and learning in online environments was exemplified best over the past year as many schools had to move their classes online all over the world (Li & Lalani, 2020). AI-enabled Adaptive Learning Systems (AI-ALS) have the potential to empower teachers and improve their efficiency with repetitive tasks such as assessment and to improve the quality of their teaching, as well as to support students in achieving their potential and improve the quality of learning (Alamri et al., 2020; Chen et al., 2020; Ghergulescu et al., 2016).

This project integrated two state-of-the-art engines, Adaptemy AI Engine[1] and AlgebraKiT Engine[2], to enhance AI-ALS with modelling sub-skills (i.e., micro-evidence within a content object like steps in a question). Modelling sub-skills enables a system to empower teachers with insights into the student's (lack of) sub-skills, to create student awareness of skills' mastery level, and to provide better learning recommendations. Mathematics is one subject where personalised learning could be improved through sub-skill modelling, as solving a maths problem is a multi-step process that requires good conceptual knowledge and procedural skills, but many students have misconceptions and gaps that can lead to errors (Hansen et al., 2020; Feldman et al., 2018).

Adaptemy AI Engine creates and updates accurate learner models and provides multi-layered adaptation and recommendations that encapsulate effective learning strategies. The effectiveness of the learning recommendations provided by the Adaptemy AI Engine was evaluated based on data from over 80k lessons (Dang & Ghergulescu, 2018). The results showed that when students followed the recommendations, they had both a higher success rate and a higher average ability improvement as compared to when the recommendations were not followed. Adaptemy AI Engine

---

[1] Adaptemy – www.adaptemy.com
[2] AlgebraKiT - https://algebrakit-learning.com/home

makes use of content and curriculum modelling to support its applicability across different courses. Furthermore, it uses machine learning to continuously update the models. Item Response Theory (IRT), Bayesian Networks, and Knowledge Space Theory are the underlying foundations of the Adaptemy AI Engine (Dang & Ghergulescu, 2018). A student response to an assessment item gives probabilistic evidence related to one concept. This evidence updates the learner model across the whole curriculum. However, Adaptemy AI Engine does not gather higher granularity evidence below concept level, which could be especially powerful for step-by-step personalised feedback and misconception detection. On the other side, AlgebraKiT Engine provides a solution that evaluates each step a student does when solving a maths problem, recognizes and explains errors automatically, and offers immediate hints to the student.

2 IMPLEMENTATION

Figure 1 presents the integration architecture of a learning platform with Adaptemy AI Engine and AlgebraKiT Engine that consists of several steps. The curriculum and content maps are created initially and taken as input by the Adaptemy AI Engine (step 1). The curriculum maps define the prerequisite relationships between the knowledge items (concepts). For example, ‘multiplication and division of integers’ is prerequisite for the knowledge item ‘order of operations’. Content objects are individual pieces of content, activities, quizzes etc., that are organised in a course structure (e.g., chapters, weeks, lessons, units), and can have metadata that reflects how they are used. Content objects may have some intended sequence and are linked to curriculum concepts.

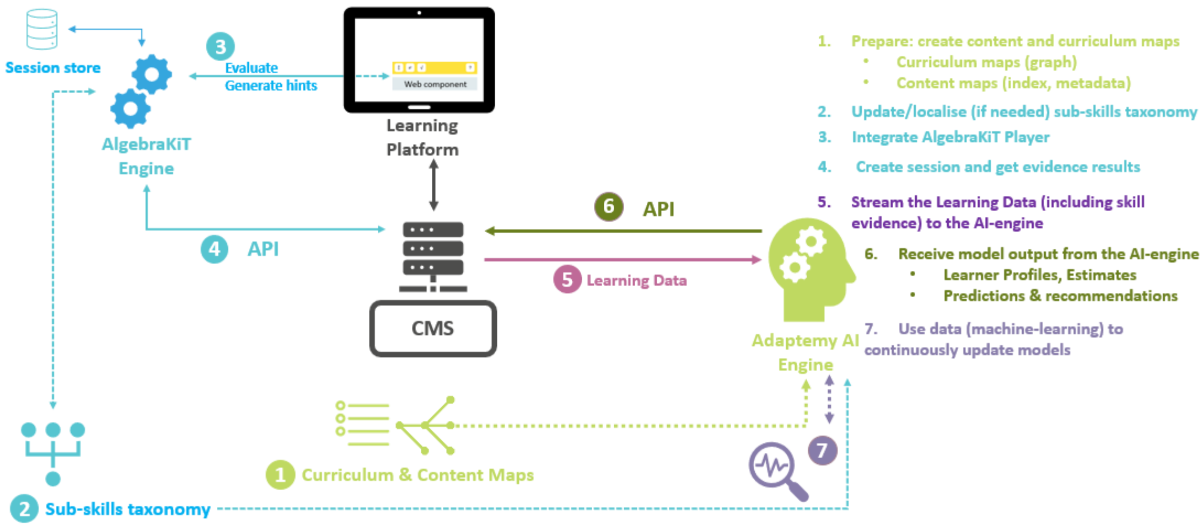


Figure 1: Integration Architecture of a Learning platform with Adaptemy’s AI Engine and AlgebraKiT Engine

The AlgebraKiT Engine integration with the learning platform consists of updating/localizing the sub-skills taxonomy if needed (step 2), integrating the AlgebraKiT Player that displays the questions and provides interactive step-by-step feedback as students practice and work through solutions (step 3), and integrating the API calls for creating sessions and getting evidence results (step 4). When a student works on a question the evidence of their maths sub-skills is retrieved through API calls by the Content Management System (CMS). The taxonomy is a hierarchical overview of the maths sub-skill tags that are recognized and derived by AlgebraKiT Engine from response analysis of student's interactions with the system. Examples of sub-skill tags for the 'Expanding Double Brackets' sub-skills collection are 'Expanding a product of two sums' and 'Expanding a square of a sum of terms'[3]. AlgebraKiT Engine can automatically collect positive and negative evidence of student's mastery of maths sub-skills. Positive evidence is generated when a student solves a problem. Negative evidence is generated when a student makes mistakes that are recognized by AlgebraKiT Engine's automatic error detection. Evidence is also generated when a student fails to solve a problem and asks for hints. In that case, AlgebraKiT Engine identifies the sub-skill related to the next maths step that the student most likely was unable to perform.

All the learning records (including evidence from AlgebraKiT) are streamed to Adaptemy AI Engine (step 5). For each student, Adaptemy AI Engine maintains an ability profile on all the concepts in the curriculum and was extended to update a profile for all the sub-skills from the taxonomy. The sub-skills are updated based on the results and evidence from AlgebraKiT Engine using a customized IRT. The CMS retrieves through API calls from the Adaptemy AI Engine the concepts profile, the sub-skills profile, student recommendations and any other learning analytics to be used in the dashboards (step 6). An example of sub-skills dashboard powered by Adaptemy AI Engine is presented in Figure 2. Adaptemy AI Engine uses student interactions data and machine learning to infer the sub-skills structure and associations between sub-skills and questions, as well as to continuously update the models (step 7). This enables Adaptemy AI Engine to make recommendations to students regarding questions they should practice to improve particular sub-skills.

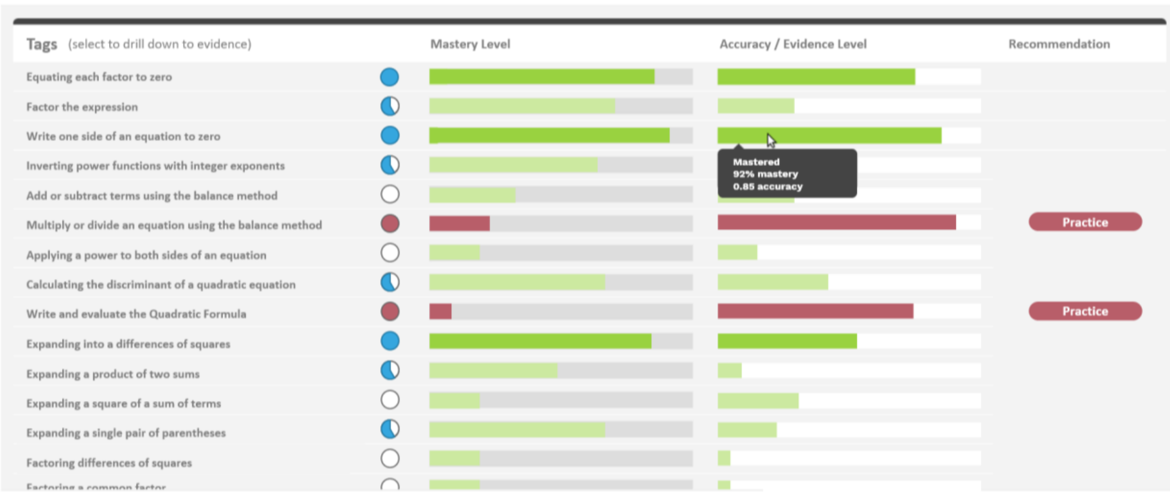


Figure 2: Sub-Skills Dashboard

3 FINDINGS

Adaptemy AI Engine and AlgebraKiT Engine were integrated with a commercial adaptive practice platform, BuildUp, that was used by 353 secondary schools in Ireland in 2020. The enhanced AI-ALS platform, BuildUp Algebra Tutor, was piloted with three classes of 5th grade students who practiced

---

[3] https://docs.algebrakit-learning.com/concepts/skill-tags/#taxonomy-of-skills

algebra questions. BuildUp Algebra Tutor was enhanced with a sub-skills learning analytics and visualization dashboard to increase students' awareness of their progress.

Before the pilot, students answered a pre-pilot survey. It is worth mentioning that 85% of students agreed that getting immediate feedback is very important in maths. 77% of students would like to use more technology in classroom when learning and practicing maths. Analyzing students' progress, it was observed that 29.3% of question workings were with at least 1 mistake and students overcame them through step-by-step feedback. The sub-skills prediction had an AUC of 0.831. The post-pilot survey assessed learner motivation (through dimensions such as interest and self-efficacy), emotions and attitudes (Bandura, 2006; Harmon-Jones, Bastian, & Harmon-Jones, 2016). Positive results between the pre- and post-survey included an increase in students' self-efficacy, interest and enthusiasm, and a decrease in students' anxiety. Students found the BuildUp Algebra Tutor easy to use, helpful in improving their maths skills, and liked practiced with the system. This paper illustrated how sub-skills modelling and learning analytics can enhance AI-ALS in real world maths courses.

REFERENCES

Alamri, H. A., Watson, S., & Watson, W. (2020). Learning Technology Models that Support Personalization within Blended Learning Environments in Higher Education. *TechTrends*, *65*(2), 62–78.

Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-Efficacy Beliefs of Adolescents*.

Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in Education: A Review. *IEEE Access*, *8*, 75264–75278.

Dang, X., & Ghergulescu, I. (2018). Effective Learning Recommendations Powered by AI Engine. *Proceeding International Workshop on Personalization Approaches in Learning Environments (PALE)*, 1-6.

Feldman, M. Q., Cho, J. Y., Ong, M., Gulwani, S., Popović, Z., & Andersen, E. (2018). Automatic Diagnosis of Students' Misconceptions in K-8 Mathematics. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–12.

Ghergulescu, I., Flynn, C., & O'Sullivan, C. (2016). *Learning Effectiveness of Adaptive Learning in Real World Context*. *EdMedia 2016-World Conference on Educational Media and Technology*, 1391–1396.

Harmon-Jones, C., Bastian, B., & Harmon-Jones, E. (2016). The Discrete Emotions Questionnaire: A New Tool for Measuring State Self-Reported Emotions. *PLoS ONE*, 11(8), e0159915.

Hansen, A., Drews, D., Dudgeon, J., Lawton, F., & Surtees, L. (2020). *Children's Errors in Mathematics* (5th ed.). London, UK: Learning Matters, SAGE.

Li, C., & Lalani, F. (2020). *The rise of online learning during the COVID-19 pandemic*. World Economic Forum. https://www.weforum.org/agenda/2020/04/coronavirus-education-global-covid19-online-digital-learning/

Estimating Ontology-based Item Difficulty Metrics for Automated Item Generation

Sean Shiverick, Clarence Dillon, Michael Smith, Jennifer Harvey
ICF International
sean.shiverick@icf.com, clarence.dillon@icf.com,
mike.smith@icf.com, jennifer.harvey@icf.com

ABSTRACT: Estimating and/or controlling the difficulty of automatically generated assessment items is necessary for producing effective learning assessments. Our team is developing a domain ontology to generate multiple response options and obtain metrics for evaluating the difficulty of generated items. Developing a comprehensive ontology can be cost prohibitive for focused training courses. To address this challenge, this project modeled the relationships between entities and concepts using network representations. Network measures (e.g., nodes, edges) provided estimates of difficulty based on salient features of the training content. The network representations may also inform ontology development.

Keywords: Assessment, Item Difficulty, Ontology, Network Representations

1 BACKGROUND

Learning assessments are used to evaluate the knowledge, skills, and abilities of learners and the effectiveness of training courses. Automated item generation (AIG) technologies can generate large numbers of assessment items that vary in quality and difficulty, which necessitates methods for controlling item difficulty. Statistical approaches are effective but require costly pre-testing. Computational models can be used to estimate difficulty for multiple-choice (MC) items based on features of the source content, item stem, response options, or by relations in a domain ontology (Kurdi et al. 2019). The more similar that incorrect distractors are to the correct response key, the more knowledge is required to distinguish them, increasing difficulty (Alsubait, et al. 2013). In an ontology, the relationships among entities, concepts, and properties are expressed as 'subject-predicate-object' triples, organized as a hierarchy, and formalized as a machine-readable graph (Vinu & Kumar, 2017). Our team is developing an ontology to generate distractors and obtain metrics by extracting semantic information from a field radio manual for an Army training course (Patten et al. 2015). A significant challenge is that the training materials do not include basic external knowledge (e.g., radios, wavelengths, etc.) needed to construct a comprehensive ontology. To address this challenge, we used network representations to estimate difficulty. In a network, words and concepts are represented by nodes and connections between concepts are called edges. Networks provide a quantitative approach for modeling cognitive structures in semantic and lexical representations (Siew et al. 2019). Networks have a graph structure that can be applied to measure difficulty.

2 IMPLEMENTATION

Four common features of item difficulty based on the relationships, connectedness, and distance in a domain ontology are shown in Table 1; each of these features corresponds to similar measures used in network representations. In considering item difficulty from a network perspective, a major assumption about memory processing is that a conceptual network is organized by semantic similarity (Siew et al. 2019). The more properties that two concepts share, the more links between the nodes, and the more closely related the concepts will be. In a memory task or during lexical retrieval, activation spreads along some number of pathways (links) in the network. The spreading activation theory (Collins & Loftus, 1975, cited in Siew et al.) predicts that retrieving one category member produces a spread of activation to other related category members. In the context of assessment, differentiating incorrect distractors that are very similar to the correct response (key) contributes to item difficulty by increasing the processing demand. Higher levels of similarity and specificity are hypothesized to increase difficulty; concepts at a deeper level in a class hierarchy are expected to yield more difficult items. By contrast, higher levels of popularity and cohesion are expected to decrease difficulty; highly connected, popular entities will produce less difficult items.

Table 1: Features of item difficulty in domain ontology and network representations.

	Domain Ontology	Network Representations
Similarity	Distance; class-relations (e.g., siblings, cousins); Jaccard similarity coefficient	Shortest path between nodes; random walk over network
Popularity	The number of connections an entity has to other entities in the ontology	Node degree or number of edges (spreading activation)
Specificity	Depth ratio: distance of an entity in a class hierarchy from the root concept	Network density; network diameter; closeness centrality
Cohesion	Cooccurrence; relative tendency of two entities to appear in the same context	Modularity; clustering; link structure; Louvain community algorithm

3 FINDINGS

Content from the training manual was converted to text and preprocessed with standard NLP tools. Neo4j software was used to construct an interactive network that supported queries about the nodes and edges to drill down into the source content. The dataset consisted of words and phrases organized hierarchically across 9 chapters (257 subsections, 2782 items) that produced a network with 7771 nodes (3498 entities, functions, procedures). General radio information in Chapter 1 (76 subsections, 257 items, 702 entities), that included sentences, bullet points, list items, figures, and tables, was represented in a network with 1098 nodes and 5127 edges (Figure 1). Node degree describes the number of connections a node has to other nodes, which corresponds to the popularity metric in an ontology. The network structure revealed a small number of highly connected nodes and large portion of nodes with few connections (Figure 2). Ten of the high degree nodes (i.e. high popularity) were: radio, screen, keys, mode, press, channel, display, user, operation, and frequency. Low degree nodes (i.e. low popularity) included: cloning, error condition, progress screens, channel parameters, waveform compatibility, configure scan, and priority channel assignment. The preliminary network measures were helpful for understanding key concepts in the domain and estimating the difficulty of distractor candidates. In an assessment, including concepts from high degree nodes may make items easier, whereas using concepts from low degree nodes may make items more difficult. Representing the training content as a network allowed us to explore salient features of the source materials that could inform ontology development. This approach may be useful for practitioners in focused training courses working with smaller ontologies.

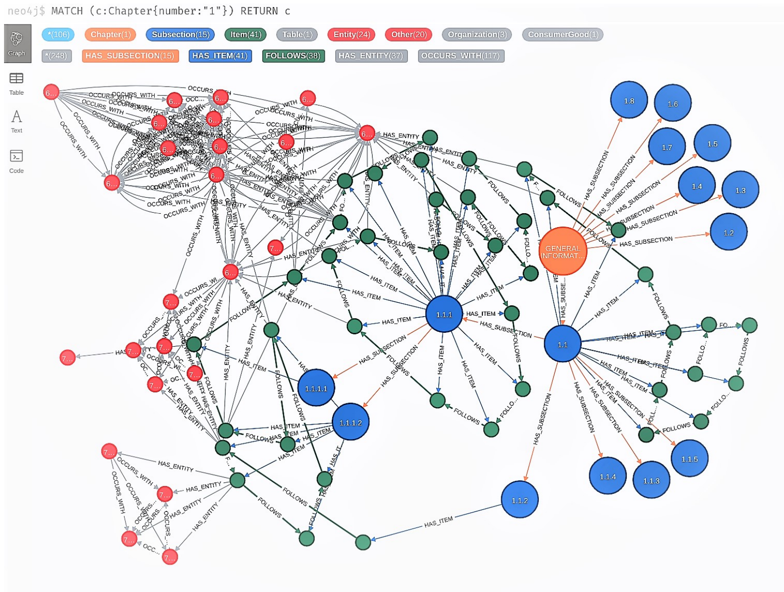


Figure 1: Network graph for chapter 1 of the field radio training manual with detail of the nodes (chapter headings, sub-headings, entities) and Edges (properties)

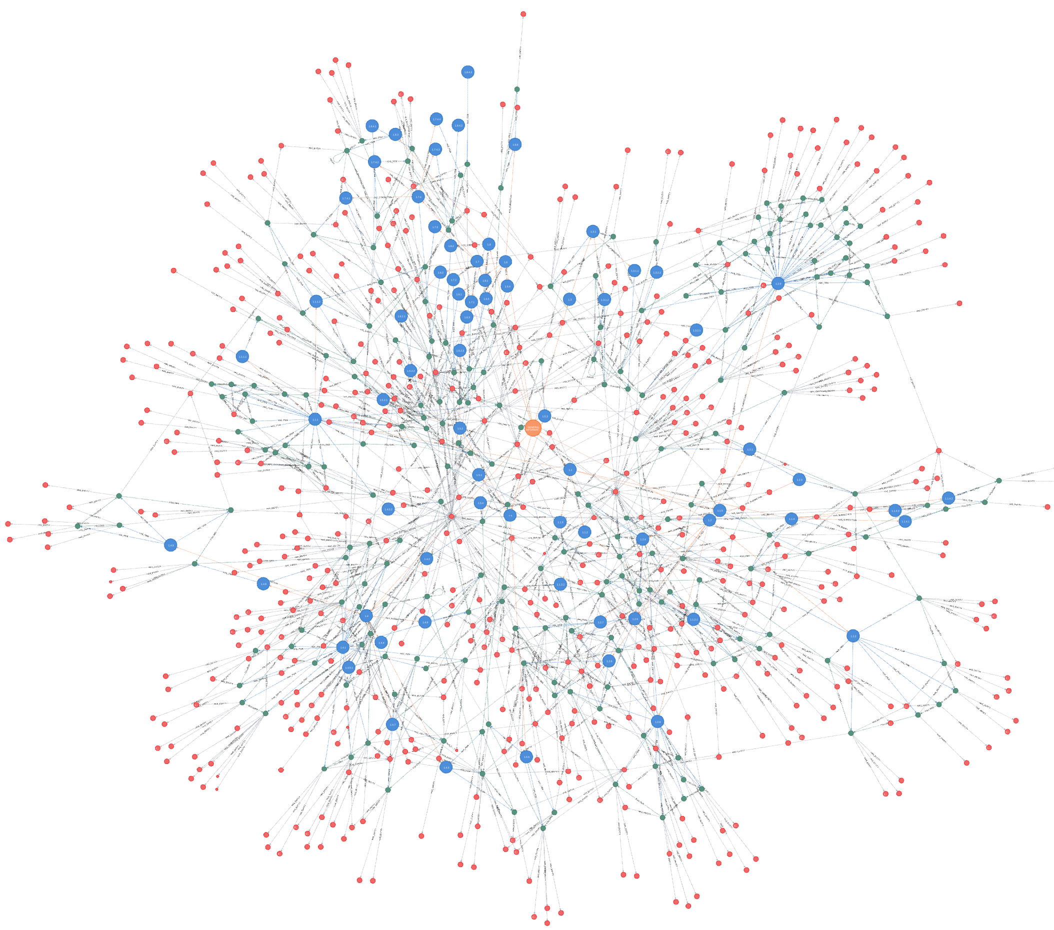


Figure 2: Expanded network graph for Chapter 1 of the field radio training manual

Authors' Note: This project conducted by ICF was supported by the Army Research Institute (ARI contract W911NF20C0018). The views, opinions, and/or findings contained in this report are those of the authors and shall not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documents.

REFERENCES

Alsubait, T., Parsia, B., & Sattler, U. (2013). A similarity-based theory of controlling mcq difficulty. *In the 2nd International Conference on e-Learning and e-Technologies in Education (ICEEE)*, pp. 283–288.

Kurdi, G., Leo, J., Matentzoglu, N., Parsia, B., Sattler, U., Al-Emari, S. (2019a). A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education, 30;* 121-204. https://doi.org/10.1007/s40593-019-00186-y.

Patten, T., Metzger, M., Hookway, S., Sliva, A., Lasser, S., Wallace, J., & Long, R. (2015). How to CREATE Domain Content for Intelligent Tutoring Systems. In the *Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium* (GIFTSym3), 125-134.

Siew, C.S.Q., Wulff, D.U., Beckage, N.M., & Kenett, Y.N. (2019). Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity,* vol. 2108423. https://doi.org/10.1155/2019/2108423.

Vinu, E.V., & Kumar, P.S. (2017). Automated generation of assessment tests from domain ontologies. *Semantic Web Journal, 8(6)*, 1023–1047. https://doi.org/10.3233/SW-170252.

The Mitigating Grade Surprise Project

Jennifer Robinson
Indiana University Bloomington
jenmetar@iu.edu

Logan Paul
Indiana University Bloomington
lopaul@iu.edu

George Rehrey
Indiana University Bloomington
grehrey@iu.edu

Aya Shohatee
Indiana University Bloomington
ahshohat@iu.edu

ABSTRACT: This paper discusses the impact grade surprise has upon students' performance, persistence and retention in five different introductory science courses. Together our courses serve approximately 6000 students per year. Data analysis indicates the existence of grade surprise and we plan to use nudges during the semester to mitigate its effect.

Keywords: Nudges, Grades, STEM, Learning Analytics,

1 BACKGROUND

Generally speaking, grade surprise (or grade anomaly) is the difference between a student's final grade in one course compared to their overall Grade Point Average (GPA) in all other courses. The presence of this grade surprise is attributed to a link between cognition and emotion, where overconfidence followed by unexpected failure makes grade surprise a painful process (Mellers et al., 2013). Analyses of grade surprise show consistent performance bias by gender (Koester et al., 2016; Matz, et al., 2017), first generation, and underrepresented minority (URM) populations (Robinson et al., 2018). However, grade surprise also presents an opportunity for students to reflect upon its causes and for teachers to intervene productively to reduce the impact of future disappointments (Robinson et al., 2018).

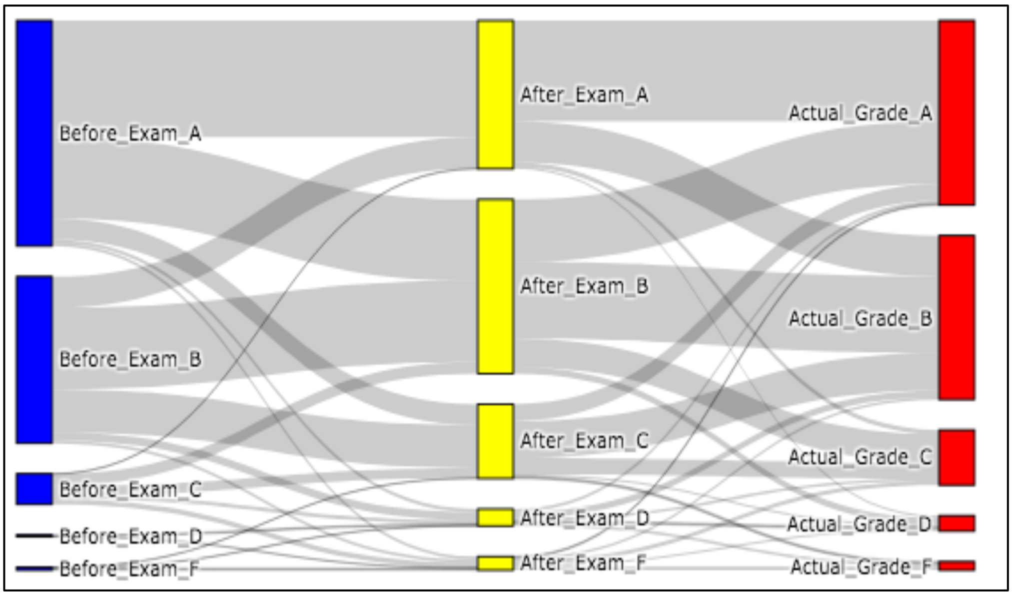
We used analytical data from our Student Information Systems (SIS) to determine the extent of grade surprise in introductory Anthropology, Biology, English, Chemistry, Informatics, and Math courses. The longitudinal data consists of academic preparation of students upon admission, academic progression such as course performance and enrollment sequences, changes in majors, demographic information, and details about retention and graduation (Rehrey et al., 2018).

2 FURTHER INVESTIGATION AND ACTIONS

Investigation into grade expectations usually compares final course grades and assumes a differential means surprise. However, our study probes the first high stakes assignment of the semester in each of the courses. A high states assignment is one that will have a substantial impact on the final course grade and is not meant as a practice assignment; this is often an exam, quiz, or test and is authentic to each individual course. A formative moment, students may experience surprise that impedes their progress and may even cause them to consider different career paths.

While SIS data dashboards did indicate that grade surprise exists in all our courses, it did not inform us of how these grades might be emotionally affecting our students. For that reason, we surveyed students about their grade expectation before the first high stakes assignment, just after completing the assignment, and compare that to their actual grade (Fig 1). Students assignment grades were then compared to their overall high school GPA, since the sample primarily consisted of freshmen and college GPA's were not available. Moreover, quantitative and open-ended survey responses were then combined with institutional and course data to understand the negative emotional consequences of grade surprise. We conducted sentiment analysis on students' open-ended survey responses to gain a better insight into has students were feeling and to categorize the results.

Figure 1: Before Exam Expectation <-> After Exam Expectation <-> Actual Grades
The letter is the response selection at each stage.



3 FINDINGS & FUTURE DIRECTIONS

Our results indicate that grade surprise does exist in all the courses analyzed. Furthermore, this discrepancy was also quite evident for some of our more at risk students, i.e.URMs, first generation students, and our 21st Century Scholars (see Figure 2 & Figure 3).

Figure 2: Grade Surprise across Courses

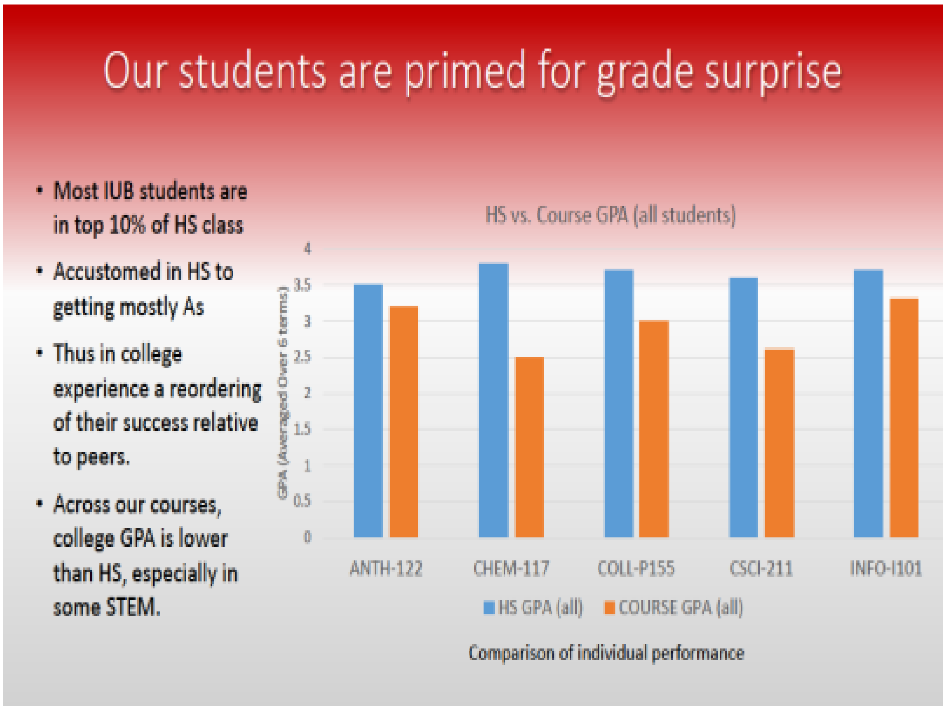
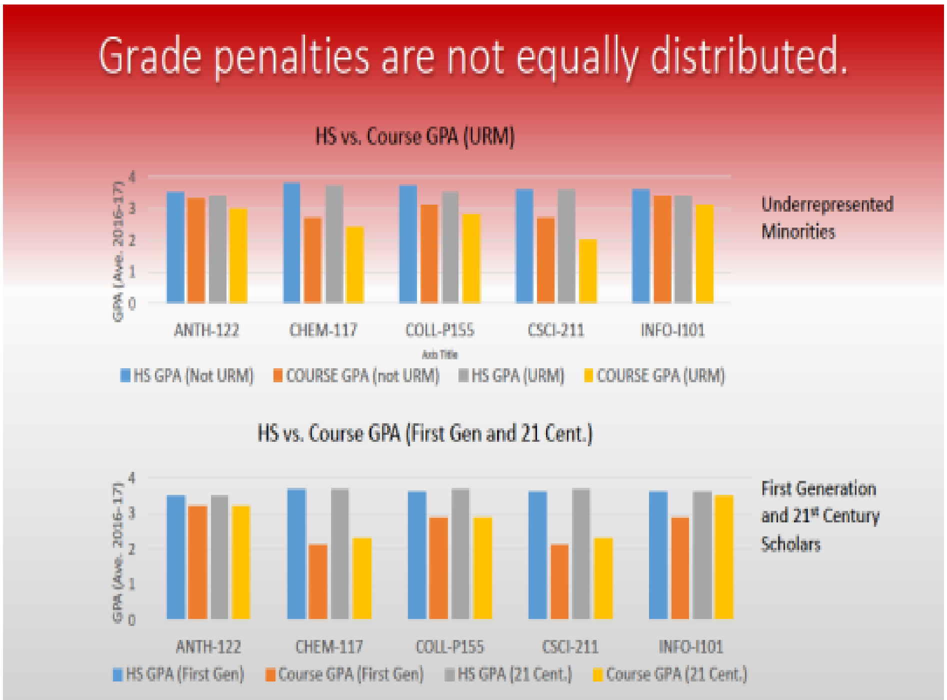


Figure 3: Grade Surprise for URM and First-Generation Students



We are currently creating a series of nudge intervention designed to mitigate grade surprise. We will be using our system-wide Student Engagement Roster (SER) to deliver personalize nudges to all our students. The nudges will be categorized and filtered based upon the student's self-reported emotional response to receiving their first grade, which will be collected immediately after their grade is posted. The SER is part of a communication network that enhances faculty/student/advisor interactions. Faculty can easily communicate with each of their students, share these messages with advisors, provide effective feedback about performance, and make recommendations. Students use the SER to monitor their academic progress in each of their enrolled courses.

The research incorporates scalability challenges from the outset. It uses institutional data and student self-reports across introductory courses in diverse disciplines, leveraging common performance behaviors and emotional responses. We recommend using the campus course management platform to scale impact across large student populations.

Our goal is to improve student success by reducing the negative effects of grade surprise that happens early in students' college career. In the next year, we plan to implement and evaluate the results of the nudge interventions and if proven to have a positive impact, make them available to 4,815 instructors to use in approximately 11,534 courses offered throughout our entire 7 campus system.

ACKNOWLEDGEMENTS

The authors of this paper would like to acknowledge that the following members of IUB's Mitigating Grade Surprise team have contributed to this study. This includes John Arthos, Amy Berndston, Jill Robinson, Chung-chieh (Ken) Shan, and Tracy Whelan. We also want to express our thanks to Indiana University's Center for Learning Analytics and Student Success (CLASS) and the Association of American Universities (AAU) for their generous support and funding of our research.

REFERENCES

Koester, B. P., Grom, G., & McKay, T. A. (2016). Patterns of gendered performance difference in introductory STEM courses.

Matz, R. L., Koester, B. P., Fiorini, S., Grom, G., Shepard, L., Stangor, C. G., Weiner, B., & McKay, T. A. (2017). Patterns of gendered performance differences in large introductory courses at five reseach universities. *American Education Research Association, 3*(4), 1-12. https://doi.org/10.1177/2332858417743754

Mellers, B., Fincher, K., Drummond, C., & Bigony, M. (2013). Surprise: A belief of an emotion? *Progress in Brain Research, 202*, 3-19. https://doi.org/10.1016/B978-0-444-62604-2.00001-0

Rehrey, G., Groth, D., Shepard, L., Fiorini, S., & Hostetter, C. (2018) Implementation of a student learning analytics fellows program. *Companion Proceedings of the 8th International Conference on Learning Analytics & Knowledge.* Sydney, Australia: Society for Learning Analytics Research.

Robinson, J., Paul, L., & Chung-chieh, S. (2018). *Understanding grade surprise: Learning analytics in first year classrooms across the curriculum*. Presented at the International Society for the Scholarship of Teaching and Learning Program, Bergen, Norway.

Guiding Principles and Strategies for Learning Analytics Implementation Focused on Equity

Kristen Fox
Tyton Partners
kfox@tytonpartners.com

Avleen Makkar
Tyton Partners
amakkar@tytonpartners.com

Justin T. Dellinger
University of Texas at Arlington
jdelling@uta.edu

ABSTRACT: In the 2019-2020 academic year, 1200 faculty and administrators at United States universities were surveyed to measure the use of learning analytics at the course level, especially those identifying disparities and driving interventions to achieve equity. In response to barriers, a set of guiding principles was developed to facilitate adoption.

Keywords: Equity, Guiding Principles, Toolkit, Case Studies, National Survey of Faculty

1 PROJECT BACKGROUND

At the end of 2019, we convened a group of experts to understand the scale and scope of barriers to the adoption of learning analytics to promote equitable learning outcomes at United States postsecondary institutions. Under pressure to reduce costs and improve services, many institutions have invested more resources in data and analytics to streamline administrative practices as well as improve instruction and learner success. Similar to the findings of the SHEILA Project, much of the work is still at an early stage in the United States. National survey data that we collected reveals that many administrators and faculty recognize the latent value in student data to promote student academic performance, and the opportunity it provides to assist institutions with implementing changes to close achievement gaps and eliminate race and income as predictors of student success. However, higher education administrators and faculty do not always have the tools to interpret learning data and make informed decisions that deliver the desired outcomes. A critical finding in our national survey is that the lack of a common set of principles and guidance are key barriers to adoption. With the support of a set of experts at various United States higher education colleges, universities, and organizations, we used the results of the survey to develop a set of guiding principles to be used by institutions seeking to implement learning analytics at scale and in service of reducing equity gaps across student groups. The full set of resources are available for download as a toolkit in the Every Learner Everywhere resource hub.

2 DESCRIPTION OF IMPLEMENTATION

Throughout the 2019-2020 academic year, we worked collaboratively with a team of institutional practitioners and experts that used the survey data to develop a toolkit to provide faculty and administrators with guidance and tools to support the adoption of learning analytics. First, we worked to develop a set of agreed- upon principles that can guide effective planning, implementation, and ongoing evaluation of learning analytics approaches. These principles focus on the following elements:

- Equity and Learning Outcomes – Explicitly set and communicate institution-level goals to achieve equity in academic outcomes across student groups through the use of learning analytics.

- Faculty, Administrator, and Student Inclusion and Support – Ensure professional development and ongoing support across stakeholders to implement, analyze, and act on data.

- Data Ethics, Privacy, and Policies – Establish and communicate institutional policies surrounding the use of student data. Policies should include fidelity and responsible use, consent and privacy, and data transparency.

- Technology and Infrastructure – Ensure that technology and infrastructure ease the ability for users to leverage student data. Outline and communicate procedures for acquiring new education technology to create a seamless integration with existing campus infrastructure.

These guiding principles have been made available to the public for widespread use and have been shared with thousands of faculty and administrators.

Second, we developed a set of institutional case studies that provide specific examples of how institutions are operationalizing and implementing these principles and a self-assessment and toolkit.

3 FINDINGS FROM PROJECT EVALUATION OR USAGE

One of the key goals of this work is to continue to share these principles with the field. We seek to help institutions implement the toolkit and use the guiding principles on their campuses. While the extent of distribution and download has been high, the extent to which institutions are applying these principles and tools is uncertain at the time of this paper given the short period of time since their release. This session seeks both to share findings and support attendees in applying them.

REFERENCES

Learning Analytics Toolkit. (2020). Every Learner Everywhere Resource Hub. Retrieved from https://solve.everylearnereverywhere.org/kit/8S6kbfaTtXefIQmP4iqq

SHEILA Project. (n.d.). Retrieved from https://sheilaproject.eu/about-sheila/

Real-time Evidence Analysis Library (REAL): Automatic Aggregation of Learning Analytics-Based Intervention

Hiroyuki Kuromiya, Taro Nakanishi
Graduate School of Informatics, Kyoto University
{Khiroyuki1993, nigiiro7716}@gmail.com

Rwitajit Maujumdar, Hiroaki Ogata
Academic Center for Computing and Media Studies, Kyoto University
{dr.rwito, hiroaki.ogata}@gmail.com

ABSTRACT: Evidence-based education has become more and more relevant in the current technology-enhanced teaching-learning era. In evidence-based education, evidence libraries play an important role in practice. In this paper, we propose a novel evidence library REAL (Real-time Evidence Analysis Library), which aims to gather and meta-analyze learning analytics-based interventions automatically. Compared to the other evidence libraries, REAL is unique in two points - 1) it has an automatic meta-analysis function of the cases, and 2) it collects cases from learning logs in real-time. Here, we show the prototype of our evidence library and discuss the potentials and limitations of such evidence library in the educational field.

Keywords: Learning Analytics, Evidence Library, Meta-Analysis, Evidence-Based Education

1 INTRODUCTION

Evidence Library plays an important role in evidence-based practice. As evidence-based education is defined as 1) utilizing existing evidence and 2) establishing sound evidence (Davies, 1999, for both points, the evidence library supports users by offering the best available cases and a place to share their original experience. In medicine, a famous evidence library, Cochran Evidence Library, is located at the center of evidence-based practice, defining the protocol of the experiment for medical researchers as well as provides reviewed evidence for practitioners (Kathie, 2002). In the educational field, What Works Clearinghouse maintained by the U.S. government arranges the results of the principal interventions in education (Schoenfeld, 2006). In the learning analytics field, LACE Evidence Hub developed by the LAK community (Ferguson & Clow, 2017) exists for archiving and sharing the effective approach in learning analytics research.

However, in the current evidence libraries, it is not assumed for practitioners to upload the evidence. In the existing libraries, evidence can only be obtained by the researchers' careful review and meta-analysis process of already published papers. It is inevitable to maintain the quality of the evidence, but at the same time, it deprives the opportunity of practitioners to share their experience. To complement the missing parts of the current evidence library systems, we propose a novel concept of evidence library, called REAL (Real-time Evidence Analysis Library), which encourages practitioners to upload and share their experience, collecting data from a learning analytics platform and aggregating multiple cases by the automatic meta-analysis. We expect that our system expands the scope of evidence-based education with a learning analytics approach.

2 OUR SOLUTION

2.1 REAL Evidence Library

To solve the problems, we developed REAL – Real-time Evidence Analysis Library. REAL is unique in the following two points. First, we expect teachers to upload their intervention results to the evidence database with a simple web form. Second, the registered cases are automatically meta-analyzed by the system so that users can see the effectiveness of the intervention at a glance. Figure 1 shows the workflow of the REAL Evidence Library. We prepare several intervention categories and users are expected to select one of the categories for registering or searching the cases.

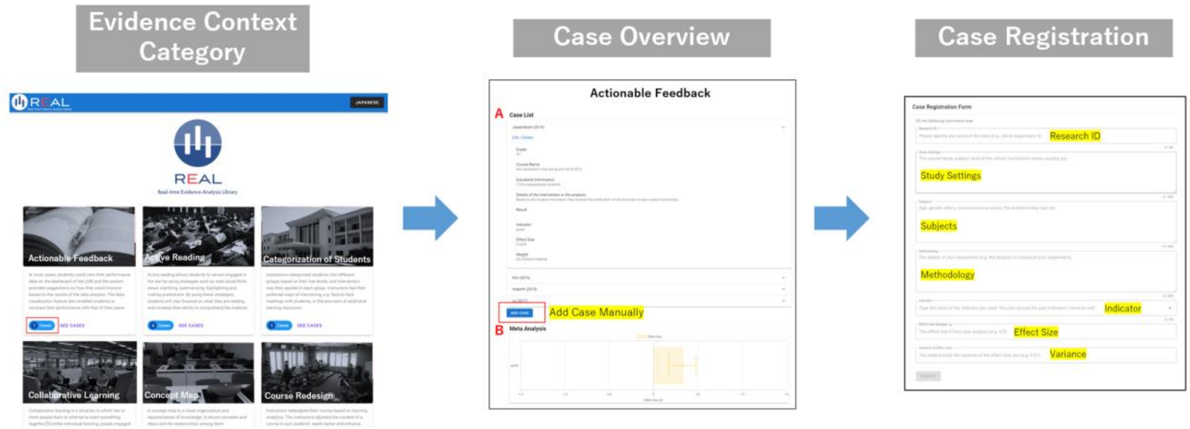


Figure 1: The Workflow of REAL Evidence Library

2.2 Automatic Meta-Analysis Process

Especially, REAL is unique in its automatic meta-analysis process. In the case overview page, users can see the unified effect size and confidence interval of the intervention to each indicator. Figure 2 shows an example of it. The length of the bar represents the unified effect size for each indicator, and the error bar represents its 95 % confidence interval. The bar chart is colored based on the popular criteria in education (Hattie, 2009). If the effect size was less than 0.2, the color will be red, yellow for greater or equal to 0.2 and less than 0.4, and blue for greater than 0.4. In the meta-analysis process, there are two popular models for integrating many cases - (1) the fixed-effect model and (2) the random-effects model. Generally speaking, the random-effects model is the extended version of the fixed-effect model. For the implementations of the automatic meta-analysis function in REAL, we compared each method in the point of the calculation cost and the range of the situations that can be handled. As a result, we decided to adopt the random effects model in REAL. The fixed-effect model is a very simple approach to meta-analyze cases, but it assumes that the variance of the effect sizes is equal in all cases. On the other hand, the random-effects model needs a bit more calculation than the fixed effect model, but it can consider the difference of the effect sizes in each case. Once users register their cases to the system, the registered case is processed and integrated with the meta-analysis instantly by the system.

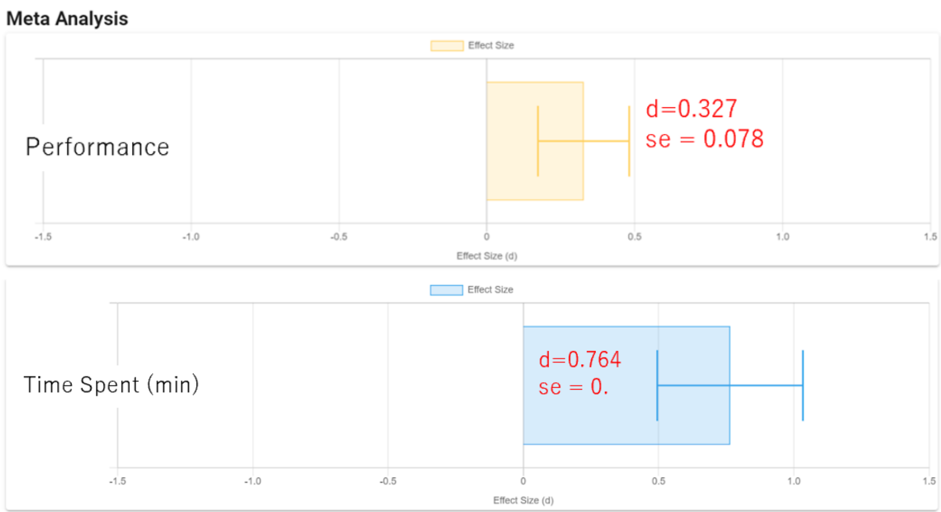


Figure 2: Real-time Meta-Analysis in REAL

3 DISCUSSION

In evidence libraries, publication bias is one of the biggest problems. In the academic culture, the study which did not show any significant results tends to be withdrawn before being reviewed by other researchers. In that context, our automatic case registration function in REAL has the potential to solve the publication bias problem because the system does not withdraw the non-significant results before registering the database. Although there are some qualitative differences in cases, our evidence library has the potential to provide an accurate estimated effect size rather than other evidence libraries in practical situations.

ACKNOWLEDGEMENT

This research was supported by JST ACT-X Grant Number JPMJAX20AA, Japan.

REFERENCES

Davies, P. (1999). What is evidence-based education? *British Journal of Educational Studies*, *47*(2), 108–121.

Ferguson, R., & Clow, D. (2017). Where is the evidence?: a call to action for learning analytics. Proceedings of the Seventh International Learning Analytics & Knowledge Conference, 56–65.

Hattie, J. (2009). *Visible Learning: A Synthesis of Over 800 Meta-Analyses Relating to Achievement*. Routledge.

Kathie Clark, A. O. (2002). The Contribution of the Cochrane Collaboration and The Canadian Cochrane Network and Centre. *The Journal of the Canadian Chiropractic Association*, *46*(3), 137.

Schoenfeld, A. H. (2006). What doesn't work: The challenge and failure of the What Works Clearinghouse to conduct meaningful reviews of studies of mathematics curricula. *Educational Researcher*, *35*(2), 13-21.

An English Picture-book Recommender System for Extensive Reading Using Vocabulary Knowledge Map

Kensuke Takii
Graduate School of Informatics, Kyoto University, Japan
kensuke.takii96@gmail.com

Brendan Flanagan
Academic Center for Computing and Media Studies, Kyoto University, Japan
flanagan.brendanjohn.4n@kyoto-u.ac.jp

Hiroaki Ogata
Academic Center for Computing and Media Studies, Kyoto University, Japan
hiroaki.ogata@gmail.com

ABSTRACT: The effectiveness of extensive reading on various linguistic skills is well understood in learning English as a foreign language. However, although the importance of personalized e-learning systems has been emphasized, few ideal recommender systems have been developed. This study proposes a recommender system of picture-books for extensive reading program which utilizes a vocabulary knowledge map. This aims for the improvement of learning efficiency of English vocabulary and the personal recommendation.

Keywords: English as a foreign language, Extensive reading, Vocabulary learning, Knowledge map, Recommendation system, Learning efficiency

1 INTRODUCTION

The methodology of extensive reading (ER) for learning English as a foreign language (EFL) and its effectiveness on various linguistic skills are well understood. So far, a lot of studies have referred to the importance of personalized e-learning recommender systems which can adapt to learners' different interests and levels. In this study, we propose a system which automatically recommends English picture-books for ER programs based on the previous activities of learners. This system mainly focuses on vocabulary learning through ER and aims at improvement of learning efficiency of EFL learning by a method based on Vygotsky's "Zone of Proximal Development (ZPD)": i.e., the system estimates each learner's English proficiency, and detects and recommends words the learner can learn efficiently. It utilizes reading logs of picture-books retrieved from an e-book reader system, BookRoll (Flanagan & Ogata, 2018), and leverages a vocabulary knowledge map (Flanagan et al., 2019) constructed with English words based on the similarity of the context in which they occur.

2 SYSTEM OVERVIEW

This system includes an e-book reader system BookRoll (Flanagan & Ogata, 2018) and a vocabulary knowledge map (Flanagan et al., 2019). As each learner's usage logs of BookRoll are recorded in Learning Record Store (LRS), we can adopt it as an interface for the ER program. A vocabulary knowledge map is a graph structure, which can be automatically generated from words in learning materials. These words are connected by an edge if they share similar contexts in which the words naturally occur. Since synonyms of words that a learner has already known are easier to learn than other words (Webb, 2007), this feature suggests that learners can learn words efficiently by learning them in succession.

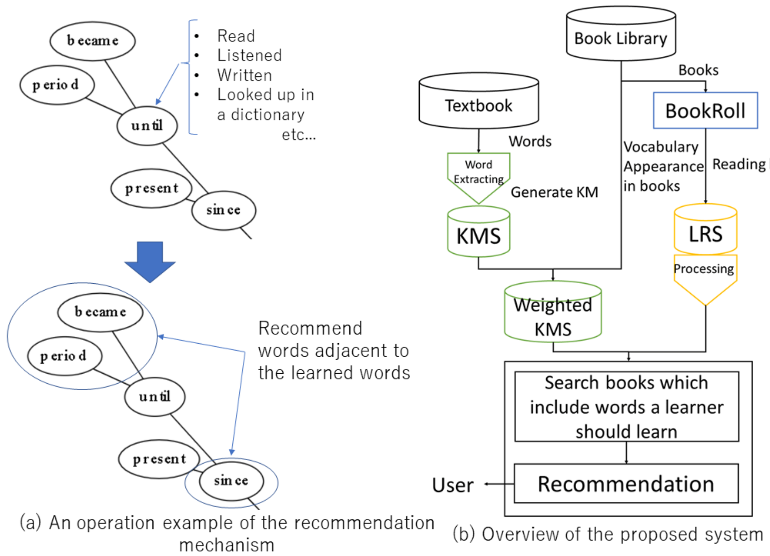


Figure 1: (a) Operation example shown with a knowledge map and (b) System overview

Figure 1(a) shows an operation example of the recommender system with a part of a vocabulary knowledge map. The words connected in the map should be learnt in succession. In this figure, if a learner learns the word "until", it is considered efficient to learn the words "became", "period" or "since" next as they share similarly contexts that are to do with "time". From this feature, the system detects words each learner has learnt and should be learn next using the learner's reading logs, and recommends picture-books which includes as many words the learner should learn as possible. Figure 1(b) shows an overview of the proposed system. First the system extracts words from a textbook and constructs a knowledge map, shown as KM in the figure, using the method proposed by Flanagan et al. (2019). The generated map is stored in the Knowledge Map Store (KMS). Then, the system extracts information on which words appear in which picture-books, links this information to the knowledge map, and stores the map in Weighted KMS. In parallel, the learners read picture-books with BookRoll, and the reading logs are stored in the LRS. If the logs include a very short browsing time, in particular 3 seconds or less per page, they are filtered and not used for recommendation. Using the remaining logs and the weighted knowledge map, the system searches books which include as many words a learner should learn as possible, and recommends them to the learner. Note that this recommendation is personalized since it is based on the learner's personal reading logs.

The user interface of the recommender system is implemented as one of the functions of LAViEW (Majumdar et al., 2019), a dashboard for analyzing learning logs retrieved from BookRoll. When a learner opens the recommendation page implemented in LAViEW, the personal recommendation for the learner is displayed. Figure 2 shows the UI design of this system.

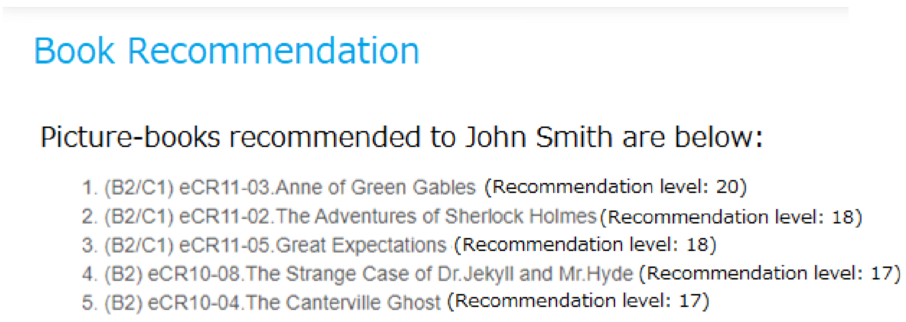


Figure 2: UI of the picture-book recommender system

In this system, 5 of the output picture-books are recommended in descending order of recommendation. The recommendation level is calculated based on the number of words which the learner should learn. When the user selects the title of a recommended e-book, they can jump to the BookRoll page of the e-book and see the word in a natural context.

3 CONCLUSION

In this study, we developed a recommender system of picture-books for ER, which aims for improvement of EFL learning efficiency by making personal recommendations which match learners' personal English proficiency. This system utilizes a vocabulary knowledge map to manage which words the learners have learnt or not. Besides, in order to track the change of the learners' English proficiency, we adopted BookRoll, whose usage logs can be collected and analyzed.

In future, we will conduct an experiment targeting 3rd graders at junior high school in Japan, and verify the effectiveness of the system we proposed in this study.

ACKNOWLEDGMENTS

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (B)20H01722, JSPS Grant-in-Aid for Scientific Research (S)16H06304 and NEDO Special Innovation Program on AI and Big Data 18102059-0.

REFERENCES

Flanagan, B., & Ogata, H. (2018). Learning analytics platform in higher education in Japan. *Knowledge Management & E-Learning: An International Journal, 10*(4), 469-484.

Flanagan, B., Chen, M. R. A., Lecailliez, L., Majumdar, R., AkçapInar, G., Ocheja, P., & Ogata, H. (2019). Automatic Vocabulary Study Map Generation by Semantic Context and Learning Material Analysis. *Proceedings of the 27th. International Conference on Computers in Education (ICCE 2019)*, 698-702

Webb, S. (2007). The effects of synonymy on second-language vocabulary learning. *Reading in a Foreign Language, 19*(2), 120-136.

Majumdar, R., Akçapınar, A., Akçapınar, G., Flanagan, B., & Ogata, H. (2019). Learning Analytics Dashboard Widgets to Author Teaching-Learning Cases for Evidence-based Education. In *Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge*. Society for Learning Analytics Research (SoLAR).

On how Unsupervised Machine Learning Can Shape Minds: a Brief Overview

Carmel Kent, Ibrahim Bashir
EDUCATE Ventures, UK
carmel@educateventures.com

Hannah Pickard, Chris Jenkins
ZISHI Adaptive, OSTC, UK

Muhammad Ali Chaudhry, Mutlu Cukurova, Rosemary Luckin
University College London

Benedict du Boulay
University of Sussex, UK

ABSTRACT: This paper briefly examines the relationship between unsupervised machine learning models, the learning affordances that such models offer, and the mental models of those who use them. We consider the unsupervised models as learning affordances. We use a case study involving unsupervised modelling via commonly used methods such as clustering, to argue that unsupervised models can be used as learning affordances, by changing participants' mental models, precisely because the models are unsupervised, and thus potentially lead to learning from unexpected or inexplicit patterns.

Keywords: Learners' mental models, unsupervised machine learning, clustering.

1 INTRODUCTION

It is well established in the learning literature that presenting learners with a simplified model of whatever is to be understood is a helpful step in learning (Seel, 2017). This paper makes the argument that machine learning (ML) models, generated by unsupervised methods, can be used as learning affordances to support and shape the development of an organization's mental models. To make that argument, we briefly outline a case study of a trading and education company (ZISHI/OSTC) which came to learn about their trainees' and mentees' behaviour via data analytics. Before using ML modelling, OSTC's trainers certainly had a strong sense that different traders traded in different ways and had developed a partial typology of trading behaviours: for example, some traders preferred to work in volatile markets, others in more stable markets. Based on such intuitions, trainers might suggest different training strategies. However, the typology had remained largely as a tacit understanding of trading behaviour. To better understand the traders' trading behaviour, we used unsupervised ML methods to arrive at four multidimensional profiles of trading behaviour. In parallel, we asked OSTC's trainers to generate their own, till then largely tacit, trading behaviour profiles into written descriptions of trading "personas". We then compared these data-driven profiles with OSTC's self-generated qualitative profiling of different kinds of traders. The data-driven profiles were then used as the predictive basis in a tool to assist OSTC to hire new traders and also formed the basis of a mentoring tool for traders currently in development.

An ML model, whether developed through supervised or unsupervised methods, will always be a simplification from a particular point of view on this complexity. This simplification and loss of detail is a strength that enables new insight; and even more so when the "point of view" on the complexity is less determined by prior expectations, such as occurs with unsupervised methods.

2 AFFORDANCES OF MACHINE LEARNING FOR HUMAN LEARNING

2.1 Human Learning

In an effective process of learning, a mental model will be stored in the long-term memory of an individual, serving later as a schema (Anderson, 1984), or a script (Preece et al., 1994). Once the model has been created, it exists independently of its sources. Visualizations, images and text can serve as mental affordances (McClelland, 2020) or as we term them – learning affordances –. These affordances may support the functionality of short-term memory (Henderson, & Tallman, 2006) to reduce cognitive load, and therefore assist learning. Our proposition is that unsupervised ML models can do that too, for example, by profiling or by simplifying and reducing the number of dimensions used.

2.2 Unsupervised Machine Learning

Raw data are not independent, contextless, self-sufficient repositories of meaning (Fjørtoft & Lai, 2020). Contextualized modelling of data, using statistical methods and, particularly ML, create possibilities for assigning existing semantics to the models, as well as for creating new semantics, which in turn, can be used as "learning affordances". The concept of affordance describes the complementary relationship between an environment and what it offers or provides to the actors within it (Gibson, 2014). The process of data modelling can start from a phase of feature engineering, in which the existing semantics can be attached to the raw data to shape it in a contextualized way. In many senses, supervised ML and reinforcement algorithms inherently include the aspiration to mimic and optimize human behaviour. Unsupervised ML, on the other hand, can reveal factors and behaviours that human guidance might have been preventing us from seeing. Unsupervised ML algorithms (such as clustering, dimension reduction or association techniques), are designed without a top-down supervision component. Thus, unsupervised algorithms are more about identification than recognition, are freer to observe the data, and are freer to learn (Amershi & Conati, 2009). In our case study. Cluster analysis was carried out and revealed four different profiles based on trading behaviour features. This was done to challenge OSTC's existing profiling mental model of traders that had been used to tailor support. We deliberately did not add to the clustered features any feature having a direct relationship with performance measures (such as profit), for the purpose of making behavioural patterns salient, and to support formative feedback.

2.3 Reflections of the Domain Experts

To explore the validity of our hypothesis that the unsupervised model had indeed affected the mental model of the organization, we invited two ZISHI/OSTC managers to compare the mental and the computed models. The interview was semi-structured around Edwards-Leis's (2012) 'transitory mental model', focusing on the model's effects on language, prediction, diagnosis and supporting their learners. In terms of the unsupervised models' affordances for human learning, it was noted that the ML model helped the trainers to focus on traders' behaviours. This contrasts with the trainers' former focus on traders' performance, which in many cases reduced to the single figure of profit. The ML models created a handy, bias-reducing shorthand to encapsulate a large number of low-level behavioural variables. These behavioural variables were usually not directly observable by the trainers themselves before the modelling, as developing such a mental model would typically take significant cognitive effort and time. In addition, the initial model was regarded as "subjective", in the sense that it had been derived from long experience of training traders, whereas the model generated with the unsupervised approach was regarded as "objective", in the sense that it had emerged from the data and was therefore trusted differently. A related difference was in the number of trading personas vs. the number of clusters. OSTC's trainers felt that they were struggling to determine what would be a sufficient set of profiles to cover the field. By contrast, arriving at four ML clusters rather than some other number was driven by the usual needs for parsimony vs. coverage of the data in unsupervised ML. Another important difference between the models was that the ML model more clearly articulated "how engaged a trader is" compared to the first model as it brought to the fore issues around order activity and diversity.

3 CONCLUSIONS

In this paper, we have very briefly described an organizational learning process, designed to help a trading and education organization develop a refined mental model of themselves via the use of unsupervised ML models. The generated model was used as a learning affordance, not just because it simplified, corrected and highlighted different aspects of an existing mental model, but also because it enabled the creation of new semantics and a new language. Using the case-study we compared the "before" and "after" models of trading behaviour. The former was subjective and formed tacitly. The latter was created via several ML methods including cluster analysis. We found that four different profiles best fitted the data, and that these had interesting similarities and differences to the "before" (subjective) version of trader personas. We acknowledge that our models were built on limited data, so future work involves remodeling as new and richer trading data become available. Further work is concentrated in designing a mentoring tool, that makes use of the profiles as the resulted profiles.

REFERENCES

Amershi, S., and Conati, C. (2009). Combining unsupervised and supervised classification to build user models for exploratory learning environments. Journal of Educational Data Mining, 1(1), 18-71.

Anderson, R. C. (1984) The notion of schemata and the educational enterprise: General discussion of the conference. In R. C Anderson, R. J. Spiro & W. E. Montague (Eds.), Schooling and the acquisition of knowledge. (1984). Hillsdale, NJ: Lawrence Erlbaum.

Edwards-Leis, C. E. (2012). Challenging learning journeys in the classroom: Using mental model theory to inform how pupils think when they are generating solutions. in T. Ginner, J. Hallstrom, & M. Hulten (Eds.), Technology Education in the 21st Century, PA.

Fjørtoft, H., and Lai, M. K. 2020. Affordances of narrative and numerical data: A social-semiotic approach to data use. Studies in Educational Evaluation, 100846.

Gibson, J. J. 2014. The ecological approach to visual perception: classic edition. Psychology Press.

Henderson, L. and Tallman, J. (2006). Stimulated recall and mental models. Lanham, ML: Scarecrow Press, Inc.

McClelland, T. (2020). The mental affordance hypothesis. Mind, 129(514), 401-427.

Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., and Carey, T. (1994). Human computer interaction, NJ: Addison Wesley.

Seel, N. M. (2017). Model-based learning: a synthesis of theory and research. Educational Technology Research and Development, 65(4), 931–966.

Learning Analytics, Performing Arts, and Teachers' Epistemic Beliefs: A Case Study of a Co-Design Process

Rebecca Nicholson[1], Colin Bone Dodds[1], Ahmed Kharrufa[1], Tom Bartindale[2]
[1]Open Lab, Newcastle University, [2] Action Lab, Monash University
r.nicholson5@newcastle.ac.uk, c.dodds2@newcastle.ac.uk,
ahmed.kharrufa@newcastle.ac.uk, tom.bartindale@monash.edu

ABSTRACT: Learning analytics should place more emphasis on adopting pedagogy-based approaches. Co-design with teachers offers one way of understanding specific disciplinary pedagogies. This paper presents a longitudinal study of a co-design process with a K-12 performing arts teacher. We found that in this case, despite a system that 'worked' in as far as the evidence from the system agreed with the teacher's own assessments, the teacher did not trust the system as it did not align with their own epistemic beliefs. Our findings suggested that more work is needed to understand teachers' epistemic beliefs within co-design processes, particularly in disciplines such as performing arts where their understanding of student achievement is not easily aligned with the quantitative data practices of learning analytics.

Keywords: Learning Analytics, Performing Arts, Epistemic Beliefs, Co-design.

1 INTRODUCTION

Recent research outlining several challenges for the adoption of learning analytics demonstrated that more emphasis needs to be placed on adopting pedagogy-based approaches to learning analytics (Tsai & Gasevic, 2017). In order to best support teachers' pedagogical practices, learning analytics tools need to account for individual contexts (Gašević, Dawson, Rogers, & Gasevic, 2016). This can vary widely as many disciplines are thought to exhibit 'signature pedagogies' (Shulman, 2005). One such example that is known to have a signature pedagogy is the performing arts which regularly set open-ended challenges, often in response to specific 'provocations' without a predetermined interpretation (Thomson, Hall, Jones, & Green, 2012), however there is a lack of research exploring the ways in which learning analytics can support pedagogy in this context.

2 METHOD / DESIGN

The design of the learning analytics tool was part of a larger embedded longitudinal research project the lead author ran alongside the Head of Performing Arts in a UK Secondary School from July 2018 to January 2020 that explored the role of technology within performing arts pedagogies. The learning analytics tool design reported in this paper was created while using project management software (basecamp.com) as an orchestration tool (Dillenbourg & Jermann, 2010) in the music classroom. During the project, the teacher wanted to use data generated from student interactions with the project management tool to inform formative assessment and feedback.

The design of the learning analytics tool took a human-centred approach (Buckingham Shum, Ferguson, & Martinez-Maldonado, 2019), drawing on the LATUX workflow (Martinez-Maldonado et al., 2015) to explore music teachers pedagogical requirements for a potential learning analytics tool.

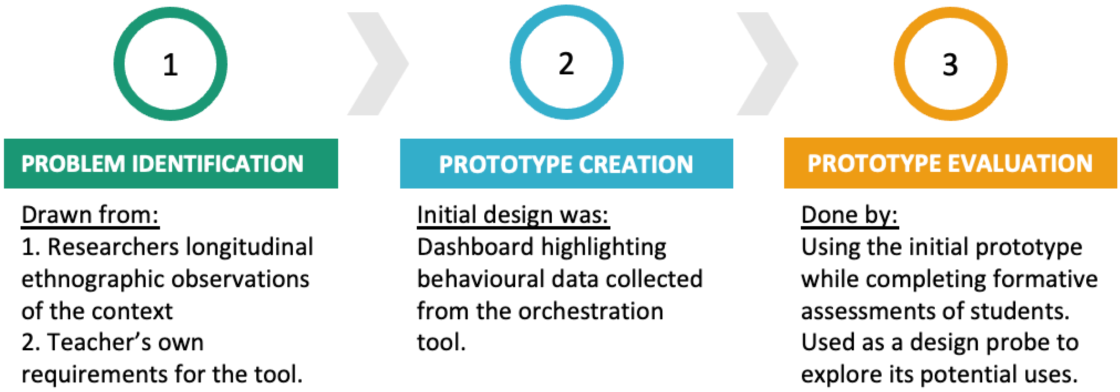


Figure 1. Design Stages following the LATUX Model

The data from the interview that formed the initial prototype evaluation was transcribed and inductively thematically analysed (Braun & Clarke, 2006). Inductive analysis was used as we sought to provide a rich narrative account of the design process and the teacher's use of the tool in a naturalistic classroom setting.

3 FINDINGS

The wider project and the use of orchestration tools in the classroom had positive feedback from the teacher and these tools have been embedded longer term into their practice. Despite the success of the overall project and the longitudinal co-design process, we found that the resulting prototype for a learning analytics system was still not likely to be utilised. The teacher felt that:

> Having ... always being very sceptical of quantitative data in music I just think I hate it, and I don't like it, and I don't trust it [Teacher].

Given the extensive co-design process and the way in which the teacher actively participated throughout this was a surprise. It was particularly unexpected given that once the teacher had spent some time with the initial prototype, exploring the data that was available for them to see they started to realise that the data from the tool matched their own assessment of pupils' progress within the classroom. It became clear that it was not a concern about the accuracy of the prototype system, but rather that their concerns lay elsewhere with the teacher saying:

> I'm looking at [Student A] and thinking it's better than anybody else's and she has probably done more and so I wish [the tool] didn't look like that because I don't want to like it, I don't want to use it.

When discussing this further it became clear that they were struggling to reconcile quantitative data practices within the learning analytics system with their own pedagogical practices telling me that

> … it's not something I've ever considered working well in music, anything statistical, you always think in terms of the qualitative.

4 IMPLICATIONS AND FUTURE WORK

The majority of existing co-design or participatory design work includes teachers in decisions once the design agenda has already been set (Prieto-Alvarez, Martinez-Maldonado, & Anderson, 2018).

Our findings suggest that research needs to engage teachers in the early stages of the co-design process to understand their epistemic beliefs before any design agendas are set. Learning analytics have been said to exist in the middle ground between epistemology, learning and assessment (Knight, Buckingham Shum, & Littleton, 2013). Knight et al. 2013 consider students' epistemic beliefs and how these are developed in relation to assessment and learning opportunities. Rather than consider students' epistemic beliefs, we suggest that understanding teachers' beliefs regarding epistemology, assessment and learning and then designing within that space could be one way in which we are able to design learning analytics that are used and embedded within classroom practices, particularly in disciplines such as the performing arts.

Future work will firstly aim to understand if these findings are replicated with other teachers of performing arts and if so will go on to explore possible co-design methodologies that seek to understand teachers' epistemic beliefs and whether this does lead to the design of learning analytics tools that are utilized in teachers' ongoing pedagogical practice.

5 REFERENCES

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology2*, *3*(2), 77–101.

Buckingham Shum, S., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-centred learning analytics. *Journal of Learning Analytics*, *6*(2), 1–9.

Dillenbourg, P., & Jermann, P. (2010). Technology for classroom orchestration. In *New science of learning* (pp. 525–552). Springer.

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *Internet and Higher Education*, *28*, 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002

Knight, S., Buckingham Shum, S., & Littleton, K. (2013). Epistemology, pedagogy, assessment and learning analytics. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 75–84).

Martinez-Maldonado, R., Pardo, A., Mirriahi, N., Yacef, K., Kay, J., & Clayphan, A. (2015). The LATUX workflow: designing and deploying awareness tools in technology-enabled learning settings. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 1–10).

Prieto-Alvarez, C. G., Martinez-Maldonado, R., & Anderson, T. (2018). Co-designing learning analytics tools with learners. *Learning Analytics in the Classroom: Translating Learning Analytics Research for Teachers*, 93–110.

Shulman, L. S. (2005). Signature pedagogies in the professions. *Daedalus*, *134*(3), 52–59.

Thomson, P., Hall, C., Jones, K., & Green, J. S. (2012). The signature pedagogies project. *The University of Nottingham: 2012*.

Tsai, Y.-S., & Gasevic, D. (2017). Learning analytics in higher education --- challenges and policies. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on - LAK '17* (pp. 233–242). https://doi.org/10.1145/3027385.3027400

Student Engagement During Remote Learning

Ethan Prihar, Anthony Botelho, Joseph Yuen, Mike Corace, Andrew Shanaj, Zekun Dai, Neil Heffernan
Worcester Polytechnic Institute
ebprihar@wpi.edu, abotelho@wpi.edu, jhyuen@wpi.edu, mlcorace@wpi.edu, ashanaj@wpi.edu, zdai@wpi.edu, nth@wpi.edu

ABSTRACT: The COVID-19 pandemic has driven a demand for transparency in online learning platforms in order to investigate the effects of remote learning on students from differing socioeconomic backgrounds. The ASSISTments learning platform has grown exponentially in users since the pandemic-induced shift to remote learning, which has provided an unprecedented opportunity to understand the effects of remote learning on groups that had not previously used online tutoring platforms. To support the learning science community, ASSISTments has compiled a comprehensive dataset on 9,609 teachers and 286,596 students who used the ASSISTments platform during the 2019-2020 school year in periods both before and after the shift to remote learning. This data was used to investigate the effects of remote learning on student engagement and revealed that teachers new to ASSISTments in low-income districts had the most difficulty maintaining student's engagement during remote learning. The full dataset is hosted by the Open Science Foundation and can be accessed at https://osf.io/q7zc5/ (Prihar, 2021).

Keywords: COVID-19, Remote Learning, Dataset, Achievement Gap

1 INTRODUCTION

The COVID-19 pandemic has had a significant impact on educational practices and policies. On March 13, 2020, the United States declared a state of emergency in response to rising COVID-19 cases, which resulted in the closure of schools across the United States. The limited resources in low-income areas prevented many students from having access to equitable educational conditions as a result of teachers needing to restructure their classes depending on available resources and access to technology (Middleton, 2020; Dewitt, 2020). In order to investigate the extent of the impact that fully-remote learning had on students in low-income school districts, and to provide the learning science community with the data to further investigate and develop methods to address this growing achievement gap, we have compiled an extensive dataset containing the complete records of the teachers and students who used the ASSISTments learning platform (Heffernan, 2014) during the 2019-2020 school year.

2 THE ASSISTMENTS 2019-2020 SCHOOL YEAR DATASET

The ASSISTments 2019-2020 school year dataset is comprised of the entirety of the interactions of students and teachers within the ASSISTments platform during the 2019-2020 school year. The dataset contains ten tables, each providing a different level of resolution for statistical analysis. The highest resolution tables provide clickstream data on students and teachers. The students' records contain information on when they took actions, e.g., answering problems, requesting tutoring, or submitting comments in the tutor. The teachers' records contain information on when they assigned homework, viewed reports on their class's progress, and how many open response questions they graded. In addition to these high-resolution action logs, the ASSISTments dataset aggregates the student action logs into problem logs, in which each log contains the details of a student completing a problem, and assignment logs, in which each log contains the details of a student completing an entire assignment. In addition to logs of teacher and student information, statistical and demographic details are provided on the students, problems, assignments, classes, teachers, and school districts that used ASSISTments during the 2019-2020 school year. A thorough dataset description can be found at https://osf.io/4nu2y/ (Prihar, 2021).

3 REGULAR-INCOME DISTRICTS VERSUS LOW-INCOME DISTRICTS

We began our analysis by investigating the effect remote learning had on student engagement in regular-income and low-income districts. Figure 1 shows the gap in average assignment completion between low-income and regular-income districts with 95% confidence bars. The difference in assignment completion between regular-income and low-income districts grew from about 4.7% to about 11.4%. This change was due to a decrease in the average assignment completion of low-income districts. Regular-income districts didn't experience a significant decrease.

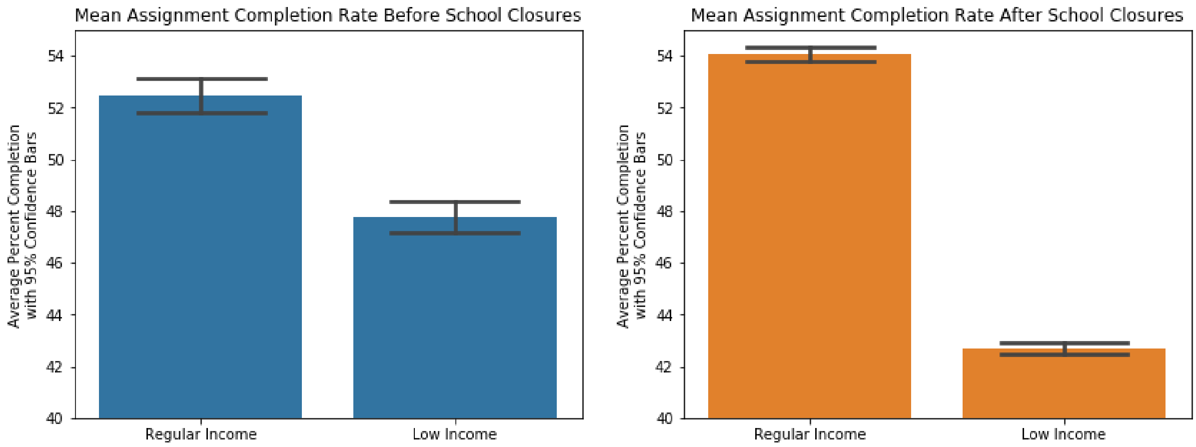


Figure 1: Average Assignment Completion in Low-Income and Regular-Income Districts Before and After the Closure

To explore the decrease in low-income students' assignment completion, the change in assignment completion was calculated separately for teachers in low-income areas who were consistently active both before and after the closure, who are referred to as persistent, and teachers from low-income areas who either started or stopped using ASSISTments after the closure, who are referred to as new. This revealed that the significant drop in low-income students' assignment completion is entirely due to new teachers. There was no statistically significant change in the assignment completion of low-income students in classes taught by persistent teachers. When looking at the difference in behavior between persistent and new teachers. We found the most significant differences were that persistent teachers viewed about 7% more reports on their class's performance on assignments, and wrote about 4% more comments to their students.

4 CONCLUSION

Although we can't directly identify the cause of high or low assignment completion percentages, this investigation led to some reassuring correlational findings. Primarily, that while most low-income student's average assignment completion fell after school closures, some teachers were able to maintain their students' engagement. The teachers who maintained their students' assignment completion did so while viewing more reports and leaving more comments than teachers whose students' assignment completion fell. This gives the impression that teachers who are more aware of their student's progress are better at keeping their students engaged.

Moving forward, the ASSISTments 2019-2020 school year dataset, and future datasets with the same format, can be used to understand the magnitude of the effect of different aspects of online instruction on student learning. The ASSISTments 2019-2020 school year dataset has potential use beyond investigating the effects of remote learning on students. The data could be used to train more robust knowledge tracing models using the skill tags associated with problems, or to create simulations of classroom environments using the student and teacher action logs. We encourage the learning science community to explore the provided data. As we receive feedback, we can improve upon the data export process and provide the learning science community with complete and open access to ASSISTments data.

ACKNOWLEDGEMENTS

We would like to thank multiple NSF grants (e.g., 1917808, 1931523, 1940236, 1917713, 1903304, 1822830, 1759229, 1724889, 1636782, 1535428, 1440753, 1316736, 1252297, 1109483, & DRL-1031398), as well as the US Department of Education for three different funding lines; a) the Institute for Education Sciences (e.g., IES R305A170137, R305A170243, R305A180401, R305A120125, R305A180401, & R305C100024), b) the Graduate Assistance in Areas of National Need program (e.g., P200A180088 & P200A150306), and c) the EIR. We also thank the Office of Naval Research (N00014-18-1-2768), Schmidt Futures, and an anonymous philanthropic foundation.

REFERENCES

Prihar, E., Botelho, A. F., Jakhmola, R., & Heffernan, N. T., III. (2021, January 25). ASSISTments 2019-2020 School Year Dataset. https://doi.org/10.17605/OSF.IO/Q7ZC5

Middleton, K. V. (2020). The Longer-Term Impact of COVID-19 on K–12 Student Learning and Assessment. *Educational Measurement: Issues and Practice*, *39*(3), 41-44. https://onlinelibrary.wiley.com/doi/full/10.1111/emip.12368

DeWitt, P. (2020). Teachers work two hours less per day during COVID-19: 8 key EdWeek survey findings. *Education Week*. https://www.edweek.org/teaching-learning/teachers-work-two-hours-less-per-day-during-covid-19-8-key-edweek-survey-findings/2020/05

Heffernan, N. T., & Heffernan, C. L. (2014). 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*(4), 470-497. https://link.springer.com/article/10.1007/s40593-014-0024-x

Dichotomous views of automation in feedback practice

Yi-Shan Tsai[1], Rafael Ferreira Mello[2], Taciana Pontual[2], Michael Burke[1], Dragan Gašević[1]
Monash University[1], Rural Federal University of Pernambuco (UFRPE)[2]
Yi-shan.tsai@monash.edu, rafael.mello@ufrpe.br, taciana.pontual@ufrpe.br,
Michael.Burke1@monash.edu, Dragan.Gasevic@monash.edu

ABSTRACT: This poster describes a pilot study at a Brazilian university, which involves a survey that seeks to explore elements of feedback that are perceived as important by students, including the role of automation. The purpose is to inform the implementation of a learning analytics-based feedback tool, OnTask. The results show that the most valued elements of feedback are the identification of attainment gaps and the relational nature of feedback. These views are also reflected in student concerns around automation, i.e., the loss of teacher-student interactions despite their positive views on efficiency and timeliness. The study concludes with the need to highlight the human inputs in the adoption of OnTask attend to the varying feedback literacy among learners.

Keywords: feedback, learning analytics, higher education, automation

1 NTRODUCTION

Feedback is a crucial part of communication between students and teachers in terms of clarifying expectations, monitoring the current progress of learners, and reflecting on the trajectory towards desired learning goals (Hattie & Timperley, 2007). Learning analytics (LA) has demonstrated great potential in enhancing feedback provision. However, the underlying pedagogies of feedback and various socio-cultural issues associated with LA are crucial to the success of LA.

As part of an initiative to adopt a LA-based feedback tool, OnTask (Pardo et al., 2019), in a Brazilian university, an online survey was conducted before OnTask was introduced, so as to identify an adoption strategy. The survey served to explore elements of feedback that are considered important by students, including the role of automation.

2 METHODOLOGY

OnTask uses rules in the form of `if this then that' to help teachers compose personalised messages based on parameters relevant to the course design (Pardo et al., 2019). Two instructors of two undergraduate courses at a Brazilian university volunteered to participate in the pilot, which took place in the first semester in 2019. The survey was sent to a total number of 60 students. In total, 36 students (31 male; 5 female) responded (response rate=58%). The respondents were aged between 17 and 47 (n=36, M=25.47, SD=7.45). The survey was designed based on prominent feedback models (Butler & Winne, 1995; Hattie & Timperley, 2007; Nicol & Macfarlane-Dick, 2006; Pardo, 2018), containing 23 questions measured by a 7-point Likert scale and 3 open-ended questions. It was developed in English (http://bit.ly/ontask_presurvey) and later translated into Brazilian Portuguese. A deliberate choice was made to administer the survey anonymously to protect student privacy, which means it would not be possible to link the responses to the academic performance of the respondents. As a result, we added a question to ask students to self-identify their performance on a scale of 1 to 10 in the course where the survey was distributed. This allows us to explore connections between self-efficacy and perceptions of feedback.

3 RESULTS

Overall, the responses show that the students were generally very positive about the feedback experience and the role of feedback in learning. The average rating scores of the questions are between 4.97 (Q23. Automation) and 6.78 (Q21. Usefulness), and the standard deviation is between 0.48 (Q21. Usefulness) and 2.10 (Q5. Connect goals) (Figure 1).

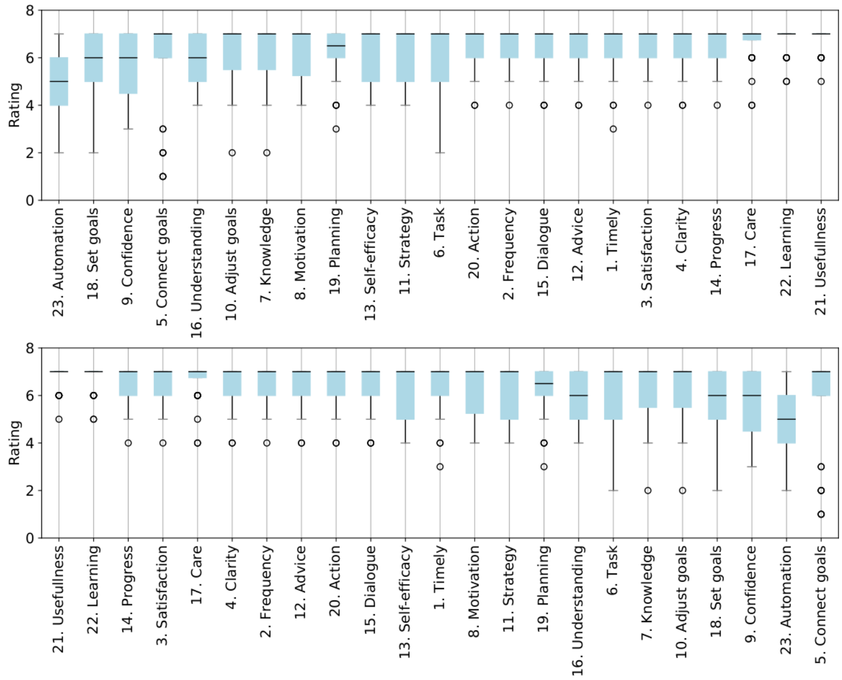


Figure 1. Box plots of responses sorted by mean (left) and variance (right)

The four statements that received the lowest average ratings with the highest variations among the responses are `I think automation can enhance the feedback process' (Q23, n=36; M=4.97; SD=1.48), `I tend to set up my own goals for course tasks' (Q18, n=36; M=5.56; SD=1.27), `The course feedback that I have received helps build my self-confidence' (Q9, n=35; M=5.57; SD=1.31), and `I can connect the course feedback that I have received with the desired goals (standards) of my course tasks' (Q5, n=36; M=5.6; SD=2.10). These findings suggest that the students were not very confident about automation in terms of enhancing feedback practice (Q23). Moreover, it appears that feedback literacy varied among the learners (Q5, Q9, and Q18).

A joint probability distribution analysis (Gaussian kernel density estimates) of the self-identified performance level and perceptions of feedback identified polarised views on topics of automation (Q23) and self-confidence (Q9) (Figure 2). The bimodal distribution shows that the extent to which existing feedback practice helps students build confidence varies even among those who self-identified as high performers. It is also notable that among this group of respondents, views on the benefits of automation are divided.

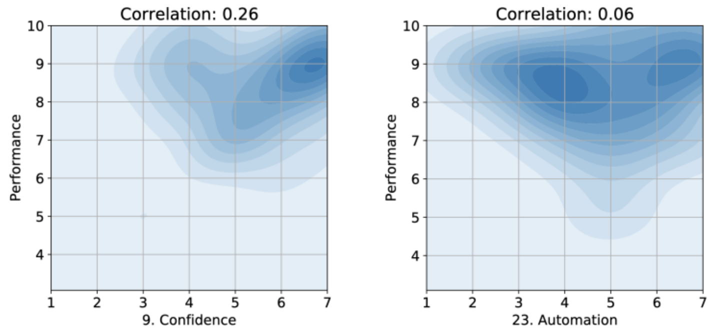


Figure 2. Polarised views on self-confidence and automation (the darker the more probable)

The responses to the three open-ended questions about the role of feedback, effective elements, and automated feedback show that students predominantly believe that the primary function of feedback is to identify strengths and weaknesses of student performance, in particular areas to improve upon and methods to achieve the desired goals. In addition, the affective dimension of feedback is particularly appreciated by the students. Sixteen respondents pointed out that feedback is important in facilitating an emotional tie between teachers and students and 7 indicated that feedback plays a key role in motivating students to pursue learning actively. Views on automation remain polarised (the `positive' code was applied 19 times and the `negative' 18 times). The three top topics are timing, relevance (personalisation) and relational. Among the positive views, economic efficiency was mentioned 14 times. Among the negative views, personalisation and relational aspects were mentioned 10 times.

4 CONCLUSION

This study shows that existing feedback experience is generally positive among the students. However, feedback to the students is not just a product, but a `relational process' that makes students feel looked after. Thus, future use of OnTask may highlight the human elements, i.e., inputs from instructors in the semi-automated process of feedback provision. Another interesting result is that variations in the perceptions of being able to connect feedback with set goals (Q5) and being able to build up self-confidence with the received feedback (Q9) indicate that feedback literacy varies among learners. Thus, when using OnTask, instructors should consider how to tailor feedback for individuals and whether further training or resources are needed to develop feedback literacy.

REFERENCES

Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, *65*(3), 245–281.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112.

Nicol, D. J., & 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.

Pardo, A. (2018). A feedback model for data-rich learning experiences. *Assessment & Evaluation in Higher Education*, *43*(3), 428–438.

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.

Seeing Spatial Reasoning

Marcelo Worsley
Northwestern University
marcelo.worsley@northwestern.edu

ABSTRACT: Spatial reasoning represents a set of skills that have broad applicability across contexts and settings. Prior research has highlighted various ways to study and measure spatial reasoning. However, many of these approaches either completely overlook the process students use to solve spatial reasoning tests or require considerable manual annotation by researchers. Motivated by work in Multimodal Learning Analytics, we propose an automated approach for extracting salient features from eye tracking data and screen recordings of students completing mental rotation tests. We test this approach with data from 19 university students. We find that several of the extracted features highly correlate with measures of student performance on the mental rotation test.

Keywords: Multimodal Learning Analytics, Gaze, Spatial Skills

1 INTRODUCTION

Spatial reasoning is generally viewed as a set of skills that have broad applicability across a variety of contexts and settings (Buckley, Seery, & Canty, 2018; Casey et al., 2008; Ramey & Uttal, 2017; Wai & Kell, 2017). When we navigate to a specific geographic location or complete a puzzle, we are employing and practicing spatial reasoning. Historically, psychologists used different psychometric tests to identify different spatial skills among research participants. However, the tests are unable to surface the ways that high and low spatial ability are evidenced in user processes. In this paper, we use eye tracking, together with computer vision and data mining, to identify process features of visual engagement that correlate with student spatial reasoning performance.

2 METHODS

Nineteen individuals participated in this study. Each student individually completed 24 mental rotation questions from a validated test bank (Ganis & Kievit, 2014). Each of the questions presented students with two objects (e.g., Figure 1). Student responses were scored for speed and correctness. Multimodal data was collected using the Social Signal Interpretation (SSI) platform (Wagner et al., 2013). SSI allows for synchronous data collection from a wide array of data streams. For this study, we collected screen recordings at five Hz, mouse tracking at 50 Hz, button presses at 50 Hz, and eye tracking at 90 Hz. A custom plugin was developed for collecting the eye tracking data. We used a Tobii 4C screen mounted eye tracker that was calibrated using 9-point calibration for each user. All other data was collected using SSI plugins that are distributed with the platform. The mental rotation test was administered as a Qualtrics survey.

Data analysis involved analysing the eye tracking data together with the screen recording videos to determine which parts of the stimuli the participants were looking at within each video frame. Achieving this output required several computer vision tasks. First, each of the original images needed to be processed to determine possible areas of interest (AOIs). Second, we processed each video to detect the location of the stimulus and translate those coordinates into absolute coordinates for the current graphic being displayed. Third, we processed the eye tracking data for fixation times and locations. We subsequently reconciled those locations with the detected AOIs. Fourth, those fixation points were used to extract features about the duration, frequency, and direction of each fixation and saccade. Finally, those features are utilized to build models and draw insights about correlations between participants' spatial reasoning and visual engagement.

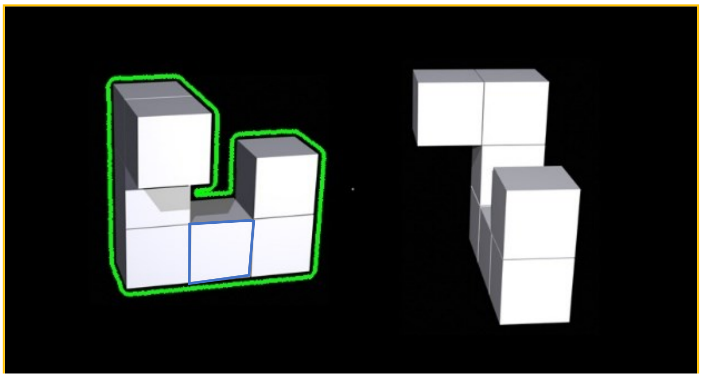


Figure 1. Sample areas of interest automatically extracted using contour detection. The green outlines one contour, while the blue outlines another, embedded contour

3 RESULTS

Table 1. Count and Percentage of Highly Correlated Features by Category and Dimension

Feature Category	Speed	Correctness	Both
AOI Fixation Duration	15 (5%)	19 (6%)	3 (1%)
AOI Fixation Count	17 (5%)	8 (3%)	2 (1%)
Mean AOI Fixation Duration	9 (3%)	45 (14%)	0 (0%)
AOI Fixation Bigrams	60 (5%)	46 (4%)	9 (1%)
Horizontal Fixation Bigrams	6 (6%)	1 (1%)	0 (0%)
Vertical Fixation Bigrams	8 (8%)	4 (4%)	3 (3%)
Targeted Structure Comparison	30 (8%)	37 (10%)	1 (0%)

There were a total of 2695 features. Each feature is associated with one of seven possible categories. In this section, we look at the percentage of each category's features that surpass the 0.5 correlation threshold relative to speed or correctness. From Table 1 we see that, despite more than 1000 AOI Fixation Bigrams, only 9% correlate with student performance. Additionally, the number of these features that correlate with speed versus correctness are comparable, suggesting that the features seem to have general utility for interrogating mental rotation performance. Targeted Region Comparison features and Total Fixation Duration features also seem to be similar across the two different dimensions. However, when we look at Horizontal Fixation Bigram features, Vertical

Fixation Bigram features, Mean AOI Fixation Duration features, and AOI Fixation Count features, we see that the number of correlating features is at least a factor of 2 difference between speed and correctness. Horizontal Fixation Bigram features, Vertical Fixation Bigram features, and AOI Fixation Count features seem to more frequently correlate with speed. In contrast, Mean AOI Fixation Duration appears to more frequently correlate with correctness than with speed.

4 DISCUSSION

This paper examines an approach for automatically generating and extracting salient features for studying mental rotation. These features were generated from a combination of eye tracking and video data and are tied to prior research on mental rotation (Just & Carpenter, 1975; Shepard & Metzler, 1971; Xu & Franconeri, 2015; Xue et al., 2017). Each of the extracted features represented one of seven categories. The goal of the analyses was to explore ways that the generated features can appropriately represent the multifaceted and complex nature of mental rotation, something that is emphasized across the vast body of prior work in this space (Buckley et al., 2018; Ramey, Stevens, & Uttal, 2018).

REFERENCES

Buckley, J., Seery, N., & Canty, D. (2018). A Heuristic Framework of Spatial Ability : a Review and Synthesis of Spatial Factor Literature to Support its Translation into STEM Education, 947–972.

Casey, B. M., Andrews, N., Schindler, H., Kersh, J. E., Samper, A., & Copley, J. (2008). The Development of Spatial Skills Through Interventions Involving Block Building Activities. *Cognition and Instruction*, *26*(3), 269–309. https://doi.org/10.1080/07370000802177177

Ganis, G., & Kievit, R. (2014, June). A new set of three-dimensional shapes for investigating mental rotation processes: validation data and stimulus set. figshare. https://doi.org/10.6084/m9.figshare.1045385.v13

Just, M. A., & Carpenter, P. A. (1975). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*(4), 441–480. https://doi.org/10.1016/0010-0285(76)90015-3

Ramey, K. E., Stevens, R., & Uttal, D. H. (2018). Steam learning in an in-school makerspace: the role of distributed spatial sensemaking. In *Proceedings of International Conference of the Learning Sciences, ICLS* (Vol. 1, pp. 168–175).

Ramey, K. E., & Uttal, D. (2017). Making Sense of Space: Distributed Spatial Sensemaking in a Middle School Summer Engineering Camp. *Journal of the Learning Sciences*, *26*(2), 277–319. https://doi.org/10.1080/10508406.2016.1277226

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703. https://doi.org/10.1126/science.171.3972.701

Wagner, J., Lingenfelser, F., Baur, T., Damian, I., Kistler, F., & André, E. (2013). The social signal interpretation (SSI) framework: multimodal signal processing and recognition in real-time. In *Proceedings of the 21st ACM international conference on Multimedia* (pp. 831–834).

Wai, J., & Kell, H. J. (2017). What Innovations Have We Already Lost?: The Importance of Identifying and Developing Spatial Talent. In M. S. Khine (Ed.), *Visual-spatial Ability in STEM Education: Transforming Research into Practice* (pp. 109–124). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-44385-0_6

Xu, Y., & Franconeri, S. L. (2015). Capacity for visual features in mental rotation. *Psychological Science*, *26*(8), 1241–1251.

Xue, J., Li, C., Quan, C., Lu, Y., Yue, J., & Zhang, C. (2017). Uncovering the cognitive processes underlying mental rotation: an eye-movement study. *Scientific Reports*, *7*(1), 10076. https://doi.org/10.1038/s41598-017-10683-6

Drinking Our Own Champagne: Analyzing the Impact of Learning-by-doing Resources in an E-learning Course

Xinying Hou, Paulo F. Carvalho, Kenneth R. Koedinger
Carnegie Mellon University
{xhou, pcarvalh, kk1u}@andrew.cmu.edu

ABSTRACT: Given the demonstrated prevalence of a “doer effect” showing that active practice is related to substantially larger learning gains than passive approaches, an important research goal is to investigate whether and how different active practice features promote students’ learning outcomes. We investigated these questions in the context of an online learning platform that teaches e-learning design principles. In particular, we considered two different practice modes - practice activities inserted in the text (inline practice) and review practice quizzes - and compared their contributions to students' learning outcomes, in terms of module quizzes, periodic exams, and course projects. Our results showed that the different practice modes had distinct impacts on learning outcomes. Doing inline practice activities contributed to students’ quiz performance at the first attempt and project performance while doing review practice quizzes helped students improve their periodic exam performance. We offer some instructional suggestions such as emphasizing practice activities that are more clearly linked with specific learning objectives for projects, and emphasizing review practice quizzes for exam preparation.

Keywords: Learning by Doing, Linear Mixed Models, Learning Outcome Prediction

1 INTRODUCTION

In the context of online learning platforms, one of the proven methods to reduce passive learning is to integrate interactive activities. A notable example is the Open Learning Initiative (OLI - Bier et al., 2014), which combines student-centered design and learning engineering approaches to provide more effective learning, such as interactive exercises embedded within the learning materials, as well as practice quizzes that allow for unlimited attempts. These activities have been shown to be more highly associated with learning outcomes than passive activities across several OLI courses, an effect known as the “doer effect” (Koedinger et al., 2015; Carvalho et al., 2017).

To extend on this prior work, it is important to further investigate how different OLI practice features contribute to learning outcomes. It is possible that some types of practice promote better learning that is more transferable, whereas other types of practice emphasize memorization and less transferable knowledge (Chi & Wylie, 2014). In this study, we compare the effect of two different types of practice activities on students’ learning outcomes. To facilitate a holistic comparison, we also considered multiple assessment measures, including module-level quizzes, periodic exams, and projects. We explore our research question in the context of an OLI course: How do different kinds of practice activities impact students’ quiz performance at the first attempt, periodic exam performance, and project performance?

2 CONTEXT AND METHODS

In this work, we analyzed students' outcome data and log data from a hybrid graduate-level course, E-Learning Design Principles and Methods. The course was hosted on the Open Learning Initiative platform, and log data from the course is hosted on the DataShop repository. Our sample consists of a total of 32 students, including 14 in Fall 2018 and 18 in Fall 2019. The course was structured into 20 modules, students in this course had "flipped homework", where they needed to learn independently and finish a graded quiz via OLI system before class meetings. In addition, students were asked to complete periodic exams (two in 2018 and four in 2019 with similar content in total), and 2 projects using knowledge taught in the course.

In OLI, each record of the student interacting with an interface element is defined as an opportunity; multiple opportunities commonly appear within a problem. To address our research questions, we defined students' count of practice as their number of opportunities on the inline activities or review practice activities. Inline practice activities are formative assessments embedded in OLI text pages. Students get immediate feedback based on their entries and could ask for hints when stuck. The content of inline questions was non-identical but matched with the content of quiz and exam questions through their alignment with specific learning goals. Beyond multiple-choice questions, in-line questions also have other formats such as drag-and-drop and matching questions. Review practice quizzes are targeted questions provided in quiz format before each periodic exam. Students get feedback once submitting a quiz attempt but no hints are delivered.

We used z-scores (i.e., standardized the score values to have zero mean and unit standard deviation) for both outcome accuracy scores and frequency of interactions, which could assist in model interpretation. For each student during each time period (e.g., between two periodic exams), we computed their review practice count, inline practice count, and total practice count. All these metrics reflect the frequency of a student's interactions with the system's practice activities. Given that students did the pre-learning quiz voluntarily and some skipped it, we used students' intercept parameter from an Additive Factors Model (AFM) provided by Data Shop and treated the normalized intercept as a measure of prior knowledge.

We began our analysis by checking whether the AFM model intercept was a good representation of students' prior knowledge. We extracted 25 students who finished more than half of the pre-learning quizzes before doing post-learning quizzes and built a Pearson correlation analysis. Our result showed a significant positive correlation between students' AFM model intercept and their average pre-learning quiz score ($r = 0.78$, $p < 0.001$), which indicates that this intercept is a good depiction of students' prior knowledge. Therefore, we use it both because it is available for all 32 students and it likely provides more information (only 9 students did all 20 pre-learning quizzes).

3 RESULTS

For each assessment, we conducted a linear mixed model with the normalized assessment score as the dependent variable; different practice count variables and student's prior knowledge as fixed-effect predictors; assessment ID as the random effect (Quiz ID, Exam ID, Project ID, respectively). Table 1 shows the final model for each assessment after the stepwise feature selection. For quiz performance, the model showed that doing inner-module inline practice activities is a significant positive predictor of quiz performance at the first attempt. For exam performance, results indicated that doing practice activities before periodic exams is a significant positive predictor of students' exam performance. For projects, students' interaction with inline practice activities is a significant positive predictor of their project performance but not for review practice activities.

Table 1: Model parameter estimates (Quiz, Periodic Exam, and Project).

Model	Parameter	Coef.	Std. Error	z
Quiz (AIC = 1628.73)	(Intercept)	-0.098	0.039	-2.524*
	inline practice count	0.144	0.038	3.817***
	prior knowledge	0.447	0.050	9.023***
Periodic Exam (AIC = 284.32)	(Intercept)	-0.066	0.099	-0.664
	total practice count	0.241	0.097	2.489*
	prior knowledge	0.309	0.131	2.348*
Project (AIC = 182.13)	(Intercept)	-0.048	0.120	-0.397
	inline practice count	0.340	0.116	2.926**
	prior knowledge	0.229	0.155	1.479

*p<0.05; **p<0.01; ***p<0.001

4 DISCUSSION AND FUTURE WORK

While prior studies have established the doer effect, our work specifically investigates how different active learning activities contribute to learning performance. A particularly novel finding is that inline practice activity is a better predictor of project performance than other forms of practice. In our context, inline practice is the one form most strongly tied to explicit learning objectives. The more successful students do inline practice questions to bolster their knowledge in ways that appear strategically connected to their needs in project-based applications. In addition, the general support for active learning, while unsurprising to many, is still helpful in addressing explicit instruction biases that many stakeholders hold, as indicated by higher effort in using, developing, or analyzing passive learning materials (e.g., lecture videos and text). Moving forward, we plan to refine the existing practice activities from both content and distribution mechanics to amplify potential impacts.

ACKNOWLEDGEMENT This work is supported by the NSF grants #1443068 and #1824257.

REFERENCES

Bier, N., Lip, S., Strader, R., Thille, C., & Zimmaro, D. (2014). An approach to knowledge component/skill modeling in online courses. *Open Learning*, 1-14.

Koedinger, K. R., Kim, J., Jia, J. Z., McLaughlin, E. A., & Bier, N. L. (2015, March). Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. In *Proceedings of the second (2015) ACM conference on learning@ scale* (pp. 111-120).

Carvalho, P. F., McLaughlin, E. A., & Koedinger, K. (2017). Is there an explicit learning bias? Students beliefs, behaviors and learning outcomes. In *CogSci*.

Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, *49*(4), 219-243.

Using enhanced Learner-facing Visual Interfaces to support Self-regulated Learning

Singh[1], S., Raković[1], M., Fan[2], Y., Lim[3], K.P., van der Graaf[4], J., Kilgour[2], K., Bannert[3], M., Molenaar[4], I., Moore[2], J., Gašević[1], D.
Monash University[1], University of Edinburgh[2]
Technical University of Munich[3], Radboud University[4]
shaveen.singh@monash.edu

ABSTRACT: Visualisations provide a rapid way for learners to see and understand their learning metrics. Yet few learner-facing interfaces have been developed to support learners' self-regulation. This paper proposes the application profile of personalised visual interfaces to support learners in self-regulated learning (SRL). Our design is theoretically based and empirically driven, and utilises trace data from multiple channels to provide clear actionable recommendations for learners to improve regulation. Guided by a quasi-experimental study in a university context, we survey the critical learning processes in SRL, describe the environment to collect multimodal and multichannel data about those processes, and suggest visualizations that can rely upon these data sources— to prompt learners to engage in metacognitive monitoring to support their regulation and learning. We conclude by outlining our next steps towards deploying and evaluating these visual interfaces in authentic learning environments to foster self-regulation to support optimal and successful learning.

Keywords: self-regulated learning; enhanced trace data; dashboards; learning analytics

1 INTRODUCTION

Providing learners with visualized information about their learning process may prompt them to reflect upon their prior and adapt their future studying, i.e., engage in metacognitive monitoring and control that are essential to productive self-regulated learning (Azevedo, Taub, & Mudrick, 2015; Roll & Winnie, 2015). Data reported in a visual form, e.g., a histogram displaying the frequency of learning strategies a student enacted over a period of observation, can cast a light on multiple elements that interplay during learning and allow researchers and educators to understand complex processes such as goal settings, enactment of learning strategies and adaptation to learning behaviours. Equally important, visualized data may afford learners the opportunity to better oversee their learning process and adapt accordingly.

We collected data from a pilot study in a university setting (n=25), where students were asked to engage in an essay writing task over the period of 45 minutes. In the task, students had to integrate three topics: Artificial Intelligence, Differentiation in the classroom and Scaffolding of learning into a 300-400 words vision essay about learning in school in 2035. The learning environment consists of six areas of interest (AOI). The AOI zones included the catalogue zone on the left, the reading and writing zones in the middle, the note taking interface (annotation tool), the planner tool, timer tools and an essay writing interface, that opens as an overlay on the screen. The choice of tools integrated and the visualisations produced were guided by the COPES model of SRL (Winne & Hadwin, 1998).

According to the COPES model, self-regulated learning spans the four phases: i) in the task definition phase, learners develop an understanding of the task, ii) during the goal setting phase, learners set their goals and plan their learning, iii) in the enactment phase, learners execute their plans and control and monitor progress iv) in the adaptation phase, adjustments are made when progress towards the goals is not proceeding as planned.

2 SRL MEASUREMENTS

We collected rich traces of these temporally unfolding SRL processes that emerged from our various data channels such as LMS log, enhanced log, eye-tracking data and interactions with external systems. Table 1 lists a subset of multichannel data sources and corresponding actions (both unobtrusive trace and self-reported data) that was captured.

Table 1: List of multi-channel data sources and their interactions that can assist self-regulation

Data Source	Event/Action
LMS Log	Content Reading, Re-reading, Content Search, Navigation Sequence Catalogue Access, Task Attempts
Enhanced Log	Mouse movements, Mouse clicks on pages, Page scroll Keyboard strokes
Eye-Tracking	Repeated number of fixations on AOIs, Sequential patterns of fixations, Revisits to AOIs, Saccades, Smooth pursuit eye movements
External Systems	**Annotation Tool:** Annotation Created, Deleted, Searched, Read **Essay Writing Tool:** Essay Write, Essay Save **Timer Tool:** Time Tracker Viewed **Planner Tool:** Planner Viewed, Planner Updated

Informed by the COPES model and the framework proposed by Siadaty, Gašević & Hatala (2016), we labeled the raw trace data into theoretically meaningful learning actions. We then interpreted the obtained patterns of learning actions as SRL processes based on our theoretical framework. These processes (Planning, Content Consumption, Working on Task, Monitoring, Evaluation) informed our design. The detailed action library and SRL labelling process can be accessed via this link.

3 FRAMING SRL SUPPORT THROUGH VISUALISATIONS: AN EXAMPLE

As highlighted earlier, our conceptual framework incorporates the COPES model that focuses on the *four phases* of SRL. Figure 1 outlines these *four phases* and the corresponding visual interfaces that are enacted to support the learner's self-regulation during the experimental task, that is to write an essay. To maintain the relevance and efficacy of SRL visualisations, we tried to coordinate the visualizations (progress indicators e.g. percentage complete) within the standard interface of the tools (such as Planner and Essay writer) for maximum potential of facilitating monitoring and regulation.

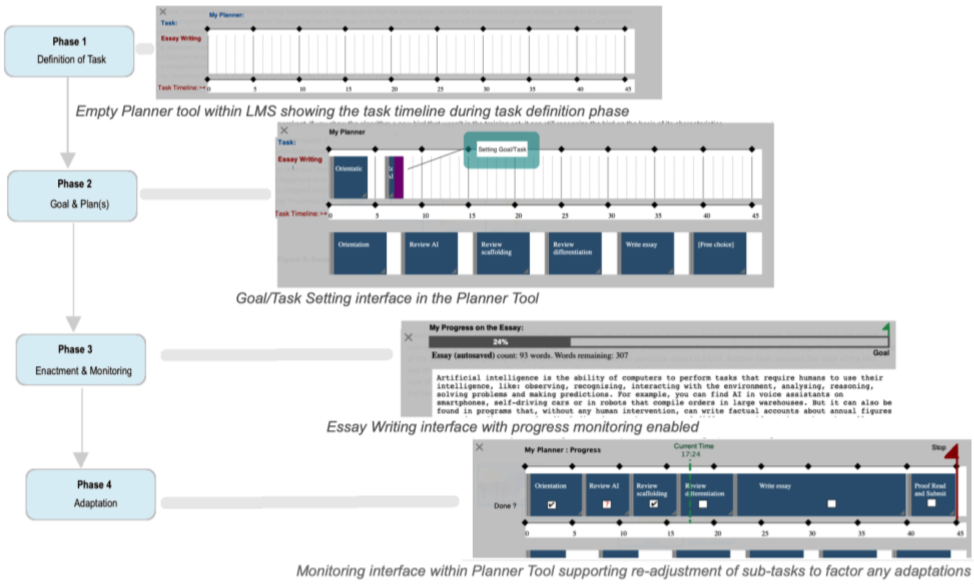


Figure 1: Interfaces enacted in line with the four phases of COPES model to support self-regulation

To support time-management, which is considered an important element in SRL (Pintrich, 2004), we captured the amount of time each learner spent on the various tools. Using the actions posited in Table 1, Figure 2 visually represents a few examples of how student's low-level interactions can be mapped onto theoretically meaningful learning actions attending to SRL criteria. Such visualisations (histogram showing SRL processes) presented in real-time can give valuable insights to learners on their current strategies with respect to time management.

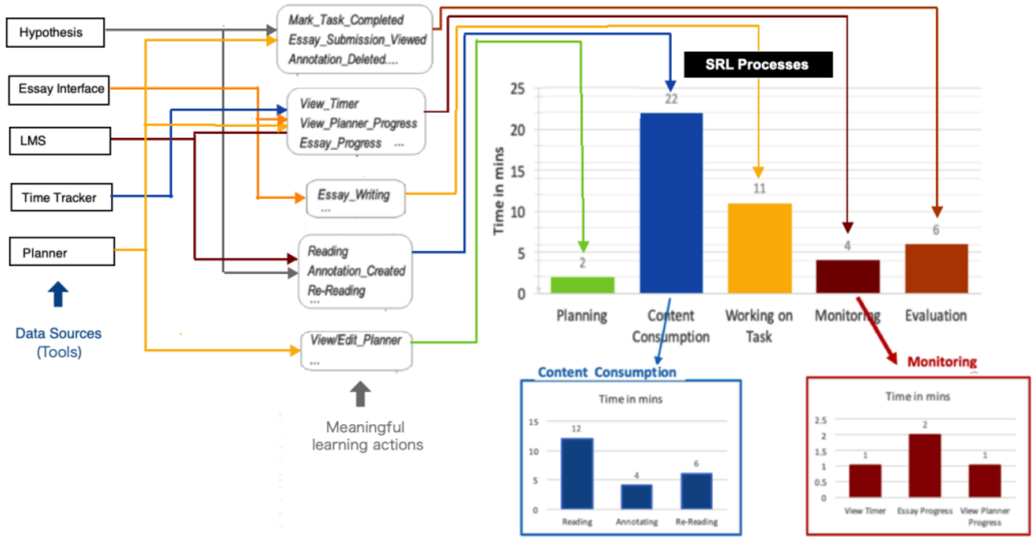


Figure 2: Visually mapping data sources and log events to specific SRL processes

4 DISCUSSION AND NEXT STEPS

Through this preliminary data analysis, we prototyped how multimodal multichannel learning data can be aligned to theoretical principles of SRL to support learners to accurately monitor and regulate their learning. An attempt has also been made to keep the balance between simple to understand and abstract monitoring indicators to maintain learner's cognitive load and reflection on one's affective reactions. Our next step is to deploy the SRL focused visual interface and catalogue of detailed visuals in an experimental study, in which students of the experimental group will have access to the personalized visualization interfaces supporting SRL and the students of the control group will not have access to these interfaces. While the current design of the learning analytics dashboard for SRL is primarily intended to be used in the laboratory setting, the overarching research program is to create the tool instrumentation (i.e., user interfaces) that can replace, to some extent, apparatus that is used in a laboratory setting (e.g., eye-trackers). This will enable our design to be used in authentic learning settings.

REFERENCES

Azevedo, R., Taub, M., & Mudrick, N. V. (2015). Technologies supporting self-regulated learning. In *The SAGE encyclopedia of educational technology* (pp. 732-735).

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review, 16*(4), 385-407.

Roll, I., & Winnie, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics, 2*(1), 7-12.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser, *Metacognition in educational theory and practice* (pp. 277-304). Mahwah, NJ, US: Lawrence Erlbaum Publishers.

Siadaty, M., Gasevic, D., & Hatala, M. (2016). *Trace-based micro-analytic measurement of self-regulated learning processes*. Journal of Learning Analytics, 3(1), 183-214.

What can we learn about college retention from student writing?

Daniel McCaffrey, Steven Holtzman, Jill Burstein & Beata Beigman Klebanov
Educational Testing Service
{dmccaffrey, sholtzman, jburstein, bbeigmanklebanov}@ets.org

ABSTRACT: Low retention rates in college is a policy concern for US postsecondary institutions, and writing is a critical competency for college (Graham, 2019). This paper describes an exploratory writing analytics study at six 4-year universities aimed at gaining insights about the relationship between college retention and writing. Findings suggest that AWE is useful for exploring the relationship between college retention and writing, and have implications for gathering diagnostic retention analytics from student writing.

Keywords: writing analytics, automated writing evaluation, higher education, retention

1 INTRODUCTION

College retention is an issue of national concern. The U.S. Department of Education, National Center for Educational Statistics (2020) reports that among first-time, full-time undergraduate students who started a 4-year Bachelor's degree in Fall 2012, only 62% completed the degree within six years – i.e., by 2018. Previous research has shown relationships between coursework writing and academic success factors. Harackiewicz et al. (2016) showed that higher utility value scores -- i.e., scores based on how a writer expresses personal relevance about technical material in a STEM writing assignment -- was correlated with STEM course retention. Using writing data from Harackiewicz et al. (2016), Beigman Klebanov et al (2017) showed that *utility value* words (e.g., our, family) were indicative of writing responses with higher human rater *utility value* scores. Allen, Dascalu, McNamara et al (2016) showed how linguistic properties in college students' writing can be used to model individual differences in students' vocabulary knowledge and reading comprehension skills. Burstein, McCaffrey, Elliot et al., (2020) used AWE to examine relationships between writing and broader academic skills and success factors (e.g., college GPA).

The study examines the question: *What relationships exist between college retention and writing*?

2 METHODS

2.1 Participants

Six four-year public universities participated in the study. One site was a Historically Black College, and a second site was a Hispanic-Serving Institution. Data from 418 students enrolled in one of the six sites were included in this study.

2.2 Data

All 418 students submitted one or more *coursework* writing assignments (n=997). Assignments were from one of these courses: first-semester English composition, Business, History, and STEM, and from argumentative, informative, or reflective genres (Burstein et al, 2019). Median coursework assignment word count was 753. A subset of 366 students completed a *timed, argumentative standardized writing assessment*; median word count was 220.

2.3 Automated writing evaluation (AWE) features

Automated writing evaluation (AWE) tools generate linguistic features from text (see Burstein et al 2017). In this study, AWE tools were used to generate 36 writing features representing six writing subconstructs: *Vocabulary* (e.g., word complexity), *English Conventions* (e.g., grammar errors), *Organization and Development* (e.g., text coherence), *Argumentation* (e.g., claim terms), *Sentence Structure* (e.g., use of clauses), and *Utility-Value language* (i.e., personal relevance terms, such as, "me", "friends"; See Beigman Klebanov, et al 2017). AWE features represent linguistic characteristics in the writing samples. To create a univariate measure for each subconstruct, the feature scores were combined into a weighted composite score. Weights equaled the loadings of the first principal component from a Principal Components Analysis fit separately for each subconstruct. Individual features were centered by genre to have mean zero. The final composite scores were standardized to a mean of zero and a variance of one and averaged across writing assignments to yield one score per composite per student. Analyses were run at the student level, and separately for the assessment and course writing data.

3 PREDICTING DROPOUT

Participating students' enrollment was tracked from 3 to 5 semesters after their participation in the study using administrative data provided by the participating universities. Random effects Cox proportional hazards regression was used to model dropout as a function of the AWE subconstruct composite score, controlling for the students' SAT/ACT score, high school GPA (HSGPA), university, and writing sample length. The models also include random effects for the course-section in which students were enrolled when participating. This accounted for possible unmodelled dropout risk factors associated with different section assignments. Separate models were fit for each feature composite score for coursework assignments, and for standardized writing assessments.

4 RESULTS

Two of the six composite features were predictive of dropout in the regression models; others were not. A standard deviation increase in the *Utility-Value language (UVL)* composite feature predicted a 26% increase in dropout hazard (i.e., dropout probability based on students continued enrollment or graduation) for both coursework ($p < 0.05$) and standardized assessment ($p < 0.10$). In addition, a standard deviation unit increase in the *Vocabulary* (VCB) composite feature in the standardized assessment predicts a 15% decrease in the hazard of dropout ($p < 0.10$). Analyses using individual component features (in the composites) showed dropout risk related positively (more risk) to pronoun use, and negatively (less risk) to use of longer words.

5 DISCUSSION

The relationship between *UVL* and college writing has not been widely studied. Beigman Klebanov et al (2017) found student success positively associated with UVL when writing assignments explicitly elicited utility value. In this study, results suggest that *UVL* could be a valuable predictor for dropout. Reviews of some student writing samples from study participants found *UVL use* reflected difficulty effectively integrating personal elements into academic writing. As discussed earlier, vocabulary has been found to be associated with various measures of academic skills and success. The results from this study extend those findings. The results suggest that exploring *vocabulary usage* with AWE might be used to identify students at risk of dropping out. More research will be required to draw clearer inferences about relationships between use of *UVL* and *VCB,* and college retention. Overall, study findings suggest relationships between AWE feature measures and retention. This insight has implications for AWE as a potential means to gather diagnostic retention analytics for stakeholders who monitor students' progress. For example, we could envision AWE integration into a learning management system in order to provide *not only* personalized learning for writing, but retention analytics for students, educators and other stakeholders to signal success and potential obstacles.

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.

REFERENCES

Allen, L., Dascalu, M., McNamara, D. S., Crossley, S., & Trausan-Matu, S. (2016). Modeling individual differences among writers using ReaderBench. In EDULearn16: 8th International Conference on Education and New Learning Technologies (pp. 5269-5279). IATED Academy.

Beigman Klebanov. B., Burstein, J., Harackiewicz, J. M., Priniski, S. J., & Mulholland, M. (2017). Reflective Writing About the Utility Value of Science as a Tool for Increasing STEM Motivation and Retention–Can AI Help Scale Up? International Journal of Artificial Intelligence in Education, 27(4), 791-818. https://doi.org/10.1007/s40593-017-0141-4

Burstein, J., McCaffrey, D., Beigman Klebanov, B., & Ling, G. (2017, September). Exploring Relationships between Writing & Broader Outcomes with Automated Writing Evaluation. In Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications (pp. 101-108). Retrieved from https://www.aclweb.org/anthology/W17-5011

Burstein, J., McCaffrey, D., Elliot, N., Beigman Klebanov, B., Molloy, H., Houghton, P. & Mladineo, Z. (2020). Exploring Writing Achievement and Genre in Postsecondary Writing . In Companion Proceedings in the 10th International Conference on Learning Analytics & Knowledge (LAK20), 53-55.

Graham, S. (2019). Changing how writing is taught. Review of Research in Education, 43(1), 277-303. https://doi.org/10.3102%2F0091732X18821125

Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Priniski, S. J., & Hyde, J. S. (2016). Closing achievement gaps with a utility-value intervention: Disentangling race and social class. Journal of personality and social psychology, 111(5), 745.

U.S. Department of Education, National Center for Education Statistics. (2020). *The Condition of Education 2020* (NCES 2020-144), Undergraduate Retention and Graduation Rates.

The country of origin as indication for cultural norms and values to personalize online courses: A recommendation for future studies

Sylvio Rüdian
Humboldt-Universität zu Berlin, Weizenbaum Institute
ruediasy@informatik.hu-berlin.de

Jana Gundlach
Universität Potsdam, Weizenbaum Institute
janagundlach@uni-potsdam.de

ABSTRACT: Adapting online courses by cultural norms and values is still underrepresented in research. An easy approach to get these norms and values is to ask for the country of origin instead of filling out a comprehensive questionnaire in a learning environment. Together with results from Hofstede, this information could theoretically be used for adapting online courses. In this paper, we show in a study with 595 participants of the US, that using the country of origin to derive norms and values at an individual level is not sufficient. The variance scores show that adapting online courses using the country-based scores should be avoided. This emphasizes the need for a questionnaire.

Keywords: Online course, adaption, personalization, culture, norms.

1 INTRODUCTION

Online courses can be adapted in different ways and personalization is required as there is no one-size-fits-all-environment. On the knowledge level, we can use tests to find knowledge gaps to create a personalized learning path. Observing a more detailed level on how knowledge should be taught based on norms and values is still undervalued in online courses. Liu et al. (2010) have shown the necessity to design online courses concerning different cultural needs to limit barriers and to ensure full participation. Wang (2006) proposed guidelines for culturally responsive online teaching. The major problem using "culture" for norms and values by countries is their mapping to individuals that can cause wrong conclusions in adaption as the tools are generally not created for an individual scale. The CVSCALE (Yoo et al., 2011) includes 26 items to get 5 norms and values at an individual level, based on the descriptions of cultural dimensions defined by Hofstede (2011). It has not to be argued that these dimensions are fruitful for learning: Long-term orientation gives insights on how learners plan to participate, the uncertainty avoidance index shows whether students need more guidance as they do not think to be confident enough or the power distance gives hints about the necessity for a hierarchy, e.g. of having a tutor and learner role. A common but insufficient solution to adapt online courses by norms and values is to use only the country of origin and the known cultural traits by Hofstede (2011) as a basis for adapting online courses. This is a low-cost approach and can be used in many existing learning environments as this information often is already existing. In this paper, we justify the need for a questionnaire to get norms and values in online learning, instead of only using the country of origin. We follow the research question: Is the country of origin sufficient to determine cultural norms and values for the adaption of online courses?

2 METHODOLOGY

We created a study with a shortened 10-item CVSCALE and asked for participants of the United States to get their answers of norms and values individually. Participants were recruited via Amazon Mechanical Turk (AMT) and received a reward of 0.60 US dollars for their participation in the study. Participants needed to have an approval rate of 95% and at least 500 completed tasks. The survey took approximately 15 minutes to complete. After screening out bots and controlling for language proficiency via an attention check, 595 participant answers were part of the final considered sample.

Answers to items of five cultural dimensions (MA: masculinity, LT: long-term orientation, CO: collectivism, UN: uncertainty avoidance, PO: power distance) were collected on behalf of Likert scales (1: strongly agree to 5: strongly disagree). Figure 1 visualizes the participants' responses, where every answer is represented by a transparent dot. The less transparent the black dot is; the more people have the cultural characteristic in common. Blue dots represent all values from Hofstede (2011), transferred to the Likert scale (1-5) for visualization.

Figure 1: Traits derived from answers by people coming from the US (country of origin).

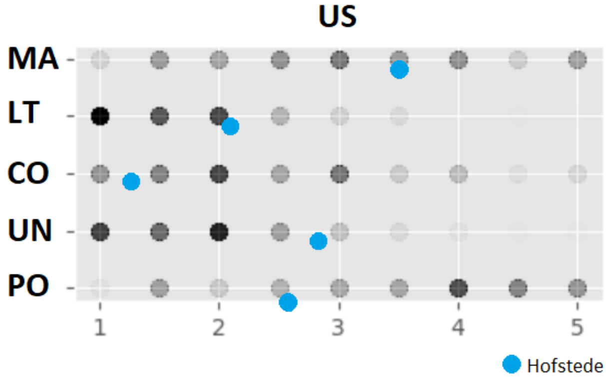


Table 1: Mean and variance of the sample.

Trait	μ	s^2
MA	3.1	1.3
LT	1.7	0.5
CO	2.4	1.0
UN	1.8	0.5
PO	3.5	1.3

The data indicate that norms and values collected on behalf of the CVSCALE do not, unlike assumed, match the cultural values linked to the countries of their origin. There is no “average user” by country and thus there cannot be an average adaption of online courses by country. The variance score s^2 in Table 1 is between 0.5 and 1.3, which is high on a scale of 1 to 5. It shows that there are diverse characteristics on the individual scale providing evidence, that within a global (online) community, cultural values merge and blur within a country. Much cross-cultural research fails to distinguish between the national or society level and the cultural one (Baskerville, 2003). We can confirm that the practice of assigning cultural country scores to individuals should be avoided in line with the findings of Taras et al. (2010). Otherwise, the adaption excludes subgroups with different cultural values that do not match with Hofstede's country scores.

3 DISCUSSION

Equating cultures with a national presence is highly problematic and does not reflect the current situation of individuals, as digitization has led to a culture and value merge. Nowadays, citizens commonly have multiple cultural and ethnic backgrounds and are in contact with individuals from various cultural backgrounds (Lee, 2010). Thereby, cultural identities play a crucial role in adopting specific cultural values (Wan et al., 2007). We, therefore, propose to regard norms and values as dimensions that are differently pronounced depending on the individual and its relative exposure. Subsequently, culture as a geographically restrained concept seems problematic and is increasingly challenged by recent developments in social anthropology (Hermans & Kempen, 1998). We limited our research to participants in the US. Using AMT to get access to participants represents a subset of the US population, but even this subset is quite diverse in norms and values. Thus, our analysis supports the argument to not use the country of origin as a general indicator for norms and values on an individual level. If we want to adapt online courses based on norms and values, we still require questionnaires to get this information. Future research can explore whether the deviation of norms and values is as diverse within the US as in other countries.

Acknowledgments: This work has been funded by the Federal Ministry of Education and Research of Germany (BMBF) under grant no. 16DII127 ("Deutsches Internet-Institut").

REFERENCES

Baskerville, R. F. (2003). Hofstede never studied culture. In *Accounting, organizations and society 28(1)*, 1-14.

Hermans, H. J., & Kempen, H. J. (1998). Moving cultures: The perilous problems of cultural dichotomies in a globalizing society. In *American psychologist(10)*, 1111.

Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. In *Online readings in psychology and culture2(1): 8*, 9.

Lee, Y.-t. (2010). Home versus host—Identifying with either, both, or neither? The relationship between dual cultural identities and intercultural effectiveness. In *International Journal of Cross Cultural Management 10(1)*, 55-76.

Liu, X., Liu, S., Lee, S.-h., & Magjuka, R. J. (2010). Cultural Differences in Online Learning: International Student Perceptions. In *Educational Technologies & Society, 13(3)*, 177-188.

Taras, V., Kirkman, B., & Steel, P. (2010). Examining the impact of Culture's Consequences: a three-decade multi-level, meta-analytic review of Hofstede's cultural value dimensions. In *Journal of Applied Psychology, Vol. 95 No. 3*, 405-439.

Wan, C., Chiu, C.-Y., Peng, S., & Tam, K.-P. (2007). Measuring cultures through intersubjective cultural norms: Implications for predicting relative identification with two or more cultures. In *Journal of Cross-Cultural Psychology38(2)*, 213-226.

Wang, M. (2006). Designing online courses that effectively engage learners from diverse cultural backgrounds. In *British Journal of Educational Technology, Vol 28(2).*

Yoo, B., Donthu, N., & Lenartowicz, T. (2011). Measuring Hofstede's five dimensions of cultural values at the individual level: Development and validation of CVSCALE. In *Journal of International Consumer Marketing 23(3-4)*, 193-210.

Developing a Framework for Mobile Assisted Language Learning through Learning Analytics for Self-Regulated Learning

Olga Viberg
KTH Royal Institute of Technology, Sweden
oviberg@kth.se

Barbara Wasson
University of Bergen, Norway
barbara.wasson@uib.no

Agnes Kukulska-Hulme
Open University, United Kingdom
agnes.kukulska-hulme@open.ac.uk

ABSTRACT: Mobile Assisted Language Learning through Learning Analytics (MALLAS) is a conceptual framework to aid learning designers in developing effective support for second- and foreign language (L2) learners through the application of learning analytics to facilitate self-regulated learning across learning settings. Designing sound support mechanisms to develop adult L2 learners' ability to self-regulate their language learning process is important since many of them have limited opportunities to participate in language classes. MALLAS can be used to assist in design choices when developing theoretically underpinned mobile assisted language learning applications and/or services.

Keywords: Mobile assisted language learning, learning analytics, self-regulated learning, learning design, support mechanisms, framework

1 INTRODUCTION & BACKGROUND

Many adult second and foreign (L2) language learners need additional support to succeed in their second language acquisition (SLA) since a common hindrance for them is that they often have insufficient opportunities to participate in language classes or lack the ability to engage in language learning on their own (Viberg, Wasson, & Kukulska-Hulme, 2020). This can be explained by the fact that many are in full time jobs or enrolled in other education. We argue that such support should focus on the development of learners' self-regulated language learning strategies, skills and knowledge that are critical for learners' ability to acquire the target language successfully (Oxford, 2016). Furthermore, we also argue that we need to carefully consider recent advancements in the fields of mobile assisted language learning (MALL; e.g., Shadiev et al., 2019;) and learning analytics (LA) for self-regulated learning (Viberg, Khalil, & Baars, 2020; Winne, 2017). Whereas the developments in these areas are frequently recognized separately, there are few efforts to draw synergy from them. Since we cannot design and improve learning directly, but only through the provision of improved conditions for L2 learners to acquire the target language effectively, based on the synergies from the aforementioned fields we offer a conceptual framework, **M**obile **A**ssisted **L**earning through **L**earning **A**nalytics for **S**elf-Regulated Learning (**MALLAS**; Viberg et al., 2020b).

MALLAS is primarily intended for learning designers to inform their design choices for mobile technology assisted support aimed at facilitating the acquisition of various language skills (e.g., speaking, and writing). This is in line with growing interest in aligning learning design and LA (Wasson & Kirschner, 2020). In the mobile learning field, such combined efforts have hitherto been rare (Pishtari et al., 2020), and we aim to fill this gap. Also, the framework contributes to the existing gap in the provision of relevant LA-grounded support mechanisms for developing learners' self-regulated learning (SRL; Viberg et al., 2020a). MALLAS is grounded in the theoretical lens of SRL (Zimmerman, 1990), strategic self-regulated language learning (Oxford, 2016), contextual mobile learning (Lincke, 2020), and the practical lens of LA. In the next section, we briefly outline MALLAS.

2 MALLAS FRAMEWORK

MALLAS (Fig.1) is a framework that captures the dimensions of self-regulated language learning (SRLL) and LA that are necessary to support MALL. It is an analytical tool that can be used to operationalise MALL support in a learning context.

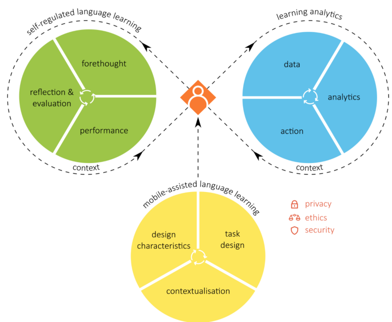


Figure 1: MALLAS

MALL has three key aspects: 1. mobile learning design characteristics, 2. contextualization, and 3. the design of language learning tasks. When developing relevant support, learning designers should consider the following design characteristics: the learner is mobile, the learner device is mobile, data services are persistent, the learning content is mobile, the learning tutor can be either an educator or an intelligent tutor (Grant, 2019; Viberg et al., 2020b). **Contextualization is** grounded in the Rich Context Model (Lincke, 2020) that includes environment, device, and personal contexts, and a MALLAS app or service will have to take these into consideration both for the data collection and recommendation services that drive the adaptivity/personalisation of the learning app or service. Language learning **task design** is supported by the task phases (i.e., forethought, performance, evaluation and reflection (S2R model of Oxford, 2016)), suggesting that language learning tasks should closely align with these SRL phases. Since SRL strategies can be taught and learnt (Viberg et al., 2020a), task design should include specific learning tasks aiming at fostering learners' SRL strategies, before they are applied to language learning. **Learning analytics** comprises *data*, *analytics*, and *action* (Fig. 1), which are used to measure and support the L2 learner's self-regulated MALL in context. Data ranges from personal (e.g., preferences) and demographic (e.g., age), location (e.g., GPS), to activity data (e.g., click stream from using an app) and data about the learning device (e.g., iPhone). This multi-channel data is used to understand the learning context, the learning path (e.g., the learner has completed all the tasks), and can be used to take action to support learning (e.g., visualisation of what a learner knows/does not know), and to recommend a learning task.

There are four main factors that will affect data quality: *richness* of the data set; *relevance* of the data; *diversity* and *quality* of the data, and *usefulness* of the findings generated by (context) analytics (Lincke, 2020). Decisions about whether to store data collected for LA locally on the learning device, or in a cloud service needs to consider privacy and security issues (Viberg et al., 2020b). Offering mixed-methods analyses, based on the theoretical lens of the S2R model (Oxford, 2016) and the examination of process-oriented behavioural data (e.g., contextual and multimodal data logs), as well as self-assessment generated data will provide a deeper understanding of the complex nature of L2 learners' SRLL processes and how support them further. The data analysis drives the adaptivity of the MALLAS app and the visualisations for the learner, educators, and researchers. The results of analytics should be used (i.e., *action)* to assist L2 learners, educators (who teach SRLL), and researchers (who help to develop relevant support tools). Overall, this poster exhibits this model of how we can harness the affordances of MALL, learning analytics, and self-regulated learning to support L2 learners through LA for SRL across learning environments.

REFERENCES

Grant, M. (2019). Difficulties in defining mobile learning: analysis, design characteristics, and implications. *Educational Technology Research and Development, 67*(2), 361-388. https://doi.org/10.1007/s11423-018-09641-4

Lincke, A. (2020). A Computational Approach for Modelling Context across Different Application Domains. Doctoral Thesis. Växjö, Linnaeus University Press. 130. http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A1421481&dswid=-5170

Oxford, R. (2016). *Teaching and researching language learning strategies: Self-regulation in context*. Second Edition. Routledge.

Pishtari, F., Rodriguez-Triana, M., Sarmiento-Marques, E., Perez-Sanagustin, M., Ruiz-Calleja, A., Santos, P., Prieto, L., Serrano-Iglesias, S., & Väljataga, T. (2020). Learning design and learning analytics in mobile and ubiquitous learning: A systematic review. *British Journal of Educational Technology, 51*(4), 1078-1100. https://doi.org/10.1111/bjet.12944

Shadiev, R., Liu, T., & Hwang, W.-Y., (2019). Review of research on mobile-assisted language learning in familiar, authentic environments. *British Journal of Educational Technology, 51*(3), 709-720. https://doi.org/10.1111/bjet.12839

Viberg, O., Khalil, M., & Baars, M. (2020a). Self-Regulated Learning and Learning Analytics in Online Learning Environments: A Review of Empirical Research. In Proceedings of the 10th International Conference on Learning Analytics and Knowledge (LAK20). https://doi.org/10.1145/3375462.3375483

Viberg, O., Wasson, B., & Kukulska-Hulme, A. (2020b). Mobile assisted language learning through learning analytics for self-regulated learning (MALLAS): A conceptual framework. *Australasian Journal of Educational Technology*, *36*(6), 34-52 https://doi.org/10.14742/ajet.6494

Wasson, B. & Kirschner, P. (2020). Learning design: European approaches. *TechTrends, 64*, 815-827. https://doi.org/10.1007/s11528-020-00498-0

Winne, P. (2017). Learning analytics for self-regulated learning. In C. Lang et al. (Eds.). *Handbook of Learning Analytics*, pp. 241-249. http://dx.doi.org/ 10.18608/hla17.021

Zimmerman, B. (1990). Self-regulated learning and academic achievement: an overview. *Educational Psychologist*, 25(1), 3-17. https://doi.org/10.1207/s15326985ep2501_2

10 Items Questionnaire for Norms and Values in Education

Sylvio Rüdian
Humboldt-Universität zu Berlin, Weizenbaum Institute
ruediasy@informatik.hu-berlin.de

Jana Gundlach
Universität Potsdam, Weizenbaum Institute
janagundlach@uni-potsdam.de

Niels Pinkwart
Humboldt-Universität zu Berlin, Weizenbaum Institute
pinkwart@hu-berlin.de

ABSTRACT: Cultural traits have still been undervalued in online courses for personalization, although there are lots of known relations concerning culture and learning. For practical experiments; there is a need to ask participants to fill in a cultural questionnaire. The CVScale can be used to collect cultural information from study participants, but the willingness to fill in the 26-item questionnaire is low due to time constraints and people tend to skip them if they are not mandatory. Thus, a questionnaire with fewer items is required for further studies concerning personalization in online courses. In this paper, we show our exploratory evaluation of a reduced 10-items questionnaire, based on the CVScale in preparation for further investigations in learning systems.

Keywords: Online course, personalization, culture, norms.

1 INTRODUCTION

Cultural traits have a strong connection to learning and yet, they are still undervalued in online courses. User modeling in adaptive online courses focuses on knowledge and competency level, and learning strategies (Matcha et al., 2020). Personalizing online courses by using norms and values at an individual level is still ignored in many practical settings. Liu et al. (2010) observed different perceptions of international students concerning online learning. Online courses need to be designed to consider different cultural needs to limit barriers and to promote participation. Wang (2006) focuses on a well-known Cultural Hofstede Dimension (Hofstede, 2011), Power Distance, proposing guidelines for culturally responsive online courses. A prominent scale which may be administered as a questionnaire is the CVScale (Yoo et al., 2011), which includes 26 items to get 5 cultural norms and values, defined by Hofstede (2011) at an individual level. With a focus on learning, having access to norms and values is insightful as they are helpful indicators for personalizing online courses by design and structure long term. It is of great interest to learn cultural traits, without the costly and time-intensive necessity to fill out a comprehensive cultural questionnaire. Rüdian et al. (2019) predict norms and values on an individual level with very high accuracy independently of questionnaires. However, this approach requires a training step at scale to create a model, which limits practical use. A shortened questionnaire to collect cultural traits may reduce dropout rates and reduces costs.

Generally, there is a tradeoff between the reduction of items within a scale and potential loss of validity. Questionnaire reduction is common: e.g. TIPI (Gosling et al., 2003) as well as the Big Five personality test from Barrick & Mount (1991), which was reduced from 50 items to 10. These shorter versions were developed for research projects that can tolerate inaccuracies. This applies to the personalization of online courses by cultural norms and values as a short cultural scale is more practical, more scalable, and less expensive.

2 METHODOLOGY & RESULTS

We tested the short scale with a questionnaire. The shortened 10 item CVScale was presented, and answers were requested on a 5-point Likert scale ranging from Strongly Agree to Strongly Disagree. We used the original CVScale (Yoo et al., 2011) as a basis and reduced this questionnaire to 10 items, two for each cultural trait (Power Distance, Long-term Orientation, Collectivism, Uncertainty Avoidance, Masculinity). The final items considered were PO4, PO5, UN2, UN3, CO3, CO4, LT4, LT6, MA2, and MA3 from the CVScale, while item selection was based on the sample used within its validation (Yoo et al., 2011) and we considered reliability and correlation. Items yielding higher Cronbach's Alphas were favored. Additionally, to capture the full breadth of a construct, selected items were chosen to be maximally different within each construct, thus covering different facets. Participants were recruited via Amazon Mechanical Turk (AMT, approval rate: 95% and HIT rate: 500) worldwide and received a reward of 0.60 US dollars. The survey took approximately 15 minutes to complete. After screening out bots and controlling for language proficiency and attention checks, 984 answers were part of the final considered sample.

Table 1: Factor Analysis (Rotated).

Components	1	2	3	4	5
PO4	.134	.331	.019	**.809**	-.122
PO5	.125	.283	-.078	**.856**	-.071
UN2	.067	-.067	**.852**	.095	.200
UN3	.096	.166	**.840**	-.119	.164
CO3	**.926**	.085	.063	.085	.078
CO4	**.911**	.075	.102	.142	.081
LT4	.117	-.09	.179	-.076	**.906**
LT6	.060	.01	.490	-.145	**.638**
MA2	.084	**.876**	.004	.275	-.051
MA3	.086	**.853**	.095	.303	-.038

The Kaiser-Meyer-Olkin (KMO) value was 0.722 (Backhaus et al., 2006) and in line with the highly significant Bartlett test, we can assume that the sample dataset is eligible for factor analysis. A rotated Confirmatory Factor Analysis (Table 1), conducted via SPSS v26, shows that all items loaded on their respective constructs and exceeded the threshold of 0.5 for qualifying as good measurements of latent constructs (Hulland, 1999). Composite reliability values exceeded the required threshold of 0.7, which is evidence for convergent validity. The reliability scores, thus the Cronbach coefficient alphas of all constructs were above the 0.7 threshold, except for Long-term Orientation (0.654). However, low-to-moderate Cronbach's alphas 0.4-0.68 are typical within short scales (Ziegler et al., 2014). The average variance extracted (AVE) and the squared roots exceeded all recommended thresholds respectively (Hulland, 1999).

OUTLOOK

We reduced the CVScale (Yoo et al., 2011), a questionnaire to obtain cultural norms and values at an individual level to use them for practical user modeling in online courses at low costs. Despite potential limitations to the AMT sample, our reduced 10 item questionnaire facilitates future studies related to the cultural adaption of online courses. It results in less accurate traits which may be outweighed by the practical benefits within the context of adapting online courses. We can use the user profile of norms and values with the reduced questionnaire to decide whether a learner wants to be guided by a tutor in an online course (PO), whether there is the need for more detailed instructions (UN), whether the learner wants to learn within a group (CO) and other derived adaptions, which are worth to investigate for optimizing the learning experience. If real psychometric values are required in detail, e.g. apart of the adaption context, we do not recommend to use the reduced CVScale.

Acknowledgments: This work has been funded by the Federal Ministry of Education and Research of Germany (BMBF) under grant no. 16DII127 ("Deutsches Internet-Institut").

REFERENCES

Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2006). Multivariate Analysemethode: Eine anwendungsorientierte Einführung. Berlin: Springer.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. In *Personnel psychology, 44(1)*, 1-26.

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the Big-Five personality domains. In *Journal of Research in personality, 37(6)*, 504-528.

Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. In *Online readings in psychology and culture2(1): 8*, 9.

Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: A review of four recent studies. In *Strategic Management Journal, 20(2)*, 195–204.

Liu, X., Liu, S., Lee, S.-h., & Magjuka, R. J. (2010). Cultural Differences in Online Learning: International Student Perceptions. In *Educational Technologies & Society, 13(3),* 177-188.

Matcha, W., Gasevic, D., Jovanovic, J., Uzir, N. A., Oliver, C., Murray, A., & Gasevic, D. (2020). Analytics of Learning Strategies: the Association with the Personality Traits. In *LAK20.*

Rüdian, S., Vladova, G., Gundlach, J., Kazimzade, G., & Pinkwart, N. (2019). Predicting culture and personality in online courses. In *Supporting Lifelong Learning (SLLL / AIED).* Chicago.

Wang, M. (2006). Designing online courses that effectively engage learners from diverse cultural backgrounds. In *British Journal of Educational Technology, Vol 28(2).*

Yoo, B., Donthu, N., & Lenartowicz, T. (2011). Measuring Hofstede's five dimensions of cultural values at the individual level: Development and validation of CVSCALE. In *Journal of International Consumer Marketing 23(3-4)*, 193-210.

Ziegler, M., Kemper, C. J., & Kruyen, P. (2014). Short scales – five misunderstandings and ways to overcome them. In *J. Individ. Dif. 35*, 185–189.

Creating a Course Recommendation System for Exchange Students

Vsevolod Suschevskiy
Centre for the Science of Learning & Technology (SLATE), University of Bergen
vsevolod.sushchevskii@uib.no

Mohammad Khalil
Centre for the Science of Learning & Technology (SLATE), University of Bergen
mohammad.khalil@uib.no

ABSTRACT: While the exchange of cross-border students in Europe has increased significantly in recent years, a growing number of these students face obstacles in selecting courses for exchange. This poster describes the first iteration of creating a course recommendation system for exchange students to select courses that fit their preferences. We implemented a combination of embedding models to enhance the course search and simplified the course selection process. Whereas the students well received the recommendation system, a grey area was found. The results of more advanced embedding models were perceived as less relevant to their expectations.

Keywords: recommendation system, higher education, data mining, exchange students

1 INTRODUCTION

The European Union has established many mobility programs across its countries to support exchanging cultural and professional experiences. However, with the volume of course-related information available to students and the nature of different course descriptions in other countries, prospect exchange students become puzzled. The wide selection of information about courses has triggered the need to help students find, organize, and use resources that match their individual goals, interests, and current knowledge (Farzan & Brusilovsky, 2006). The poster at hand presents the first iteration of CERES, a Course Recommendation System for Exchange Students. CERES employs data science models, including natural language processing techniques (e.g., Universal Sentence Encoder and information retrieval) to support the University of Bergen in helping exchange students select courses based on available course catalogues. CERES also offers features such as filtration and a course shopping cart to facilitate course selection and management (see Figure 1, right).

A growing number of course recommendation systems are recognized as forms of learning analytics interventions (Khalil & Ebner, 2015). Many of these were built in recent years (Bodily & Verbert, 2017) and for regular students (e.g., goal-based models, Jiang et al., 2019). However, there is scarce research on recommendation systems dedicated for exchange students for multiple reasons: restricted access to respondents, the complex interlinkage to several institutions of exchange agreements, and conflicting students' goals, i.e. academic and recreational (Badstübner & Ecke, 2009). In our work, we also faced data scarcity on exchange students, which acted as a bottleneck to building enough knowledge for the course recommendation system. Having these challenges in mind, we designed CERES to support exchange students, better understand their needs, and collect data about their course choice behavior using quantitative data (log data, pre-and post-survey) and qualitative data, including virtual interviews for a more optimized course recommendation system.

2 CERES- STRUCTURE & DESIGN

Our recommendation system uses two main algorithms to retrieve matching courses: I) Universal Sentence Encoder, a model that represents documents as vectors and II) TF-IDF, an information retrieval model that measures words relevance of a corpus. For the Universal Sentence Encoder, A typical interaction scenario with CERES starts in a browser, where a student deals with a user-friendly Shiny R web application. Each search query the student enters is sent to a dedicated R application via Plumber API, where words are transformed into embeddings — a vector representing the meaning of a word in a numeric form. In this way, we dramatically increased the application's performance, so those parallel sessions will not interfere with each other. We use Universal Sentence Encoder (see Cer et al., 2018) implementation in TensorFlow. That vector is compared to vectors of course descriptions stored in Elasticsearch, then a list of similar courses is returned to the student. Further, all logs are sent to Elasticsearch, where they are stored for learning analytics exploration (see Figure 1, left).

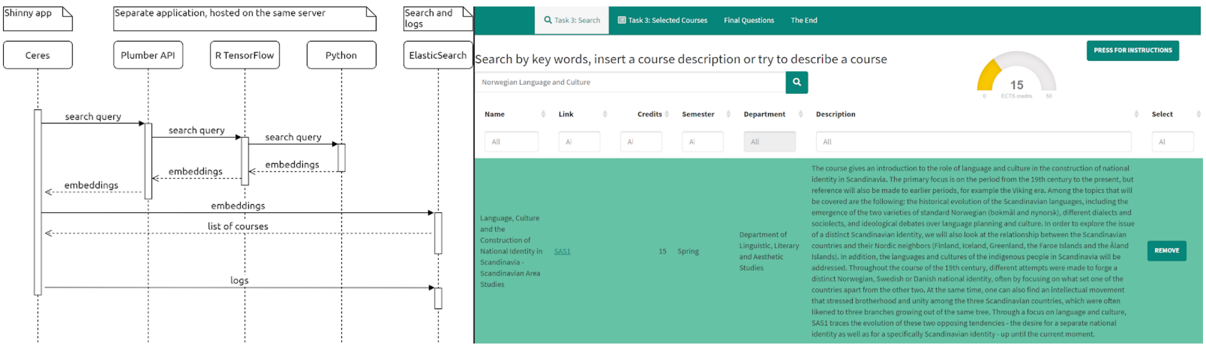


Figure 1: CERES architecture (left); CERES user interface (right).

For the TF-IDF model, we use Elasticsearch that looks for an exact match of a query and course description catalogue. CERES selects either one of these algorithms based on a simple rule: If the input query is longer than two words or the query has words that do not appear in the description, the system will use Universal Sentence Encoder to show partly relevant or relevant results, otherwise, it will return exact matching using the TF-IDF model.

3 EVALUATION

We wanted to evaluate CERES in terms of the quality of the returned results and its usability. The pre-and post-survey questionnaires were designed to understand students' motives and demographics. A link to a survey and experimenting CERES to rate relevant and irrelevant courses were posted by the University's international office on Facebook groups to recruit exchange students. Participants were placed in a draw for a 10 EUR worth of gift cards. In total, we had 34 students who finished the survey, seven of whom volunteered to participate in a 30-minute long semi-structured virtual interview.

The preliminary results from the survey and CERES were positively evaluated by the exchange students. Over 80% of them strongly agree that CERES will be useful for future exchange students. During the interviews, some students experienced course recommendations that depart from what they expect from the system. This is, in fact, commonly usual according to Pardos and Jiang (2020). One of the students stated that “if I am just like inserting keywords, then I would expect it to show me courses that match the same keywords”.

4 FUTURE WORK

We believe there is room for more improvements in terms of the study design and refinement of the used algorithms. We are setting up a large-scale user testing study to have more generalizable results and polish CERES before the launch. Thus, we are conducting an experiment to see the difference between the two models’ perceptions. A more complex model might be less exact for exchange students who must overcome several additional problems. Considering we want to support an exploratory search; we need to introduce the system in more detail and provide more support for a search or roll back to a more interpretable exact matching model without unexpectedness and serendipity.

5 ACKNOWLEDGEMENTS

This research work is supported by the Peder Sather Grant Program under No. 11331.

REFERENCES

Badstübner, T., & Ecke, P. (2009). Student expectations, motivations, target language use, and perceived learning progress in a summer study abroad program in Germany. Die Unterrichtspraxis/Teaching German, 42(1), 41-49.

Bodily, R., & Verbert, K. (2017). Review of Research on Student-Facing Learning Analytics Dashboards and Educational Recommender Systems. IEEE Transactions on Learning Technologies, 10(4), 405–418. https://doi.org/10.1109/TLT.2017.2740172

Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., …, & Kurzweil, R. (2018). Universal Sentence Encoder. ArXiv:1803.11175 [Cs]. http://arxiv.org/abs/1803.11175

Farzan, R., & Brusilovsky, P. (2006). Social Navigation Support in a Course Recommendation System. In V. P. Wade, H. Ashman, & B. Smyth (Eds.), Adaptive Hypermedia and Adaptive Web-Based Systems (pp. 91–100). Springer. https://doi.org/10.1007/11768012_11

Jiang, W., Pardos, Z. A., & Wei, Q. (2019). Goal-based Course Recommendation. Proceedings of the 9th International Conference on Learning Analytics & Knowledge, 36–45. https://doi.org/10.1145/3303772.3303814

Khalil, M., & Ebner, M. (2015). Learning analytics: principles and constraints. In *EdMedia+ Innovate Learning* (pp. 1789-1799). Association for the Advancement of Computing in Education (AACE).

Pardos, Z. A., & Jiang, W. (2020). Designing for serendipity in a university course recommendation system. In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge (pp. 350-359).

Enabling Multimodal Reading Analytics through GOAL Platform

Rwitajit Majumdar, Duygu Şahin, Taisho Kondo,
Huiyong Li, Yuanyuan Yang, Brendan Flanagan and Hiroaki Ogata
Kyoto University
dr.rwito@gmail.com

ABSTRACT: Multi-modal analytics has the potential for understanding learning activities and possibly supporting the process based on the interpretation of the signals captured. However, the integration of multiple data sources remains an issue. While some commercial packages assist researchers to organize the data for their specific study, the infrastructure is still not available to integrate learning logs and physiological sensor data together for the same learning episode. In our prior work, GOAL system solves the issue of data integration by connecting sensor data from wearable activity trackers by API and linking the data to the learners registered in a learning management system through LTI. In this work, we extend the functions to collect EEG, GSR and eye tracking data from physiological sensors within the same technical infrastructure. A pilot study was conducted to synchronize data during a reading-based learning task and discuss the capabilities of such a platform for designing learning support at scale within our learning and evidence analytics framework (LEAF).

Keywords: Multimodal Analytics; GOAL; Smart watch; EEG; GSR; Eye-tracking

1 PRACTICAL CHALLENGES OF MULTIMODAL DATA INTEGRATION

Technology enhanced learning systems aim to support a lifelong learning agenda of 21st century learners. In that regard, any learning experience also depends on learners' physiological and psychological state leading to effective learning engagement. Off the shelf physical activity trackers provide daily activity levels of the learners such as steps taken, sleep patterns and pulse rate, etc. In addition to that, it is possible to collect high resolution physiological data related with the variation of the electrical properties of the skin by electrodermal activity (EDA) sensors and the electrical activity of the brain by electroencephalography (EEG) sensors. Furthermore, eye-tracking devices can capture the eye gaze data of the user. Such physiological and physical behavior data can be processed to provide indicators related to various cognitive states of the learners during any learning episode. This is the field of multi-modal learning, which has progressed technologically by integrating computational methods to process the captured signals and theoretical models to make sense of how learning is happening. However, this process of knowledge building through MMLA faces various ethical, practical and methodological challenges (Cukurova et al. 2020). One of the practical challenges is to synchronize data from various sensors and link it to the learner's learning activity logs. Currently there exists no platform that provides such data synthesis.

Here we present our technical platform GOAL (Majumdar et al.2018), whose functionalities were extended to link learning logs and physiological sensor data of learners. We demonstrate it with an actual data collection session and discuss possibilities of such multi-source data.

2 USING GOAL FOR DATA INTEGRATION OF MULTIMODAL DATA

Earlier we developed GOAL, a system to integrate the learning logs with learners' physical activity data from smart watches and mobile applications. Reading interactions such as navigation (change of page) and annotation (markers and memos) were the learning logs captured from BookRoll (Ogata et al. 2015), an eBook-based learning platform that is linked to the Learning Management System through Learning Tool Interoperability (LTI) standards. Similarly, GOAL can also be linked to the LMS through LTI. It synchronizes the learners' physical activity data from Google, Apple and Garmin devices through APIs. The data is linked to the unique user identification (UUID) which is generated from the LMS and thereby pseudonymized and recorded in the Learning Record Store (LRS). We use the same technical architecture to integrate the data from physiological sensors. The data pipeline captures heart rate, EDA and temperature data from Empatica E4 devices and EEG data from Emotiv devices. A data upload option caters to any sensors that provide CSV files with time information. The overview of the technical architecture is provided in Figure 1a.

An authentic reading-based learning context was designed and the basic data synchronization pipeline was tested with a pilot study. The sequence of learner activity and the initial data collection process is presented in Figure 1b. Participants were asked to click on the EmpaticaE4's button after each activity was completed to gather the temporal markers of the start and the completion of events.

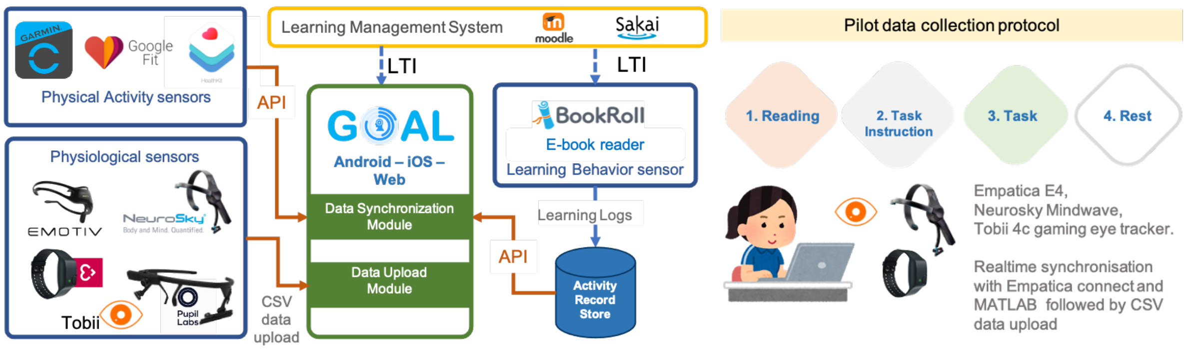


Figure 1: a. Technical architecture to synchronize multi-modal data in GOAL b. Pilot protocol

3 RESULTS AND DISCUSSION

Figure 2 below provides the visualized processed data collected from one participant during the pilot activity. It highlights the synchronized timeline marked by the participant's use of empaticaE4's data marker button. The BookRoll logs on the same timeline provided an additional channel of reading interaction data. Based on the EEG signals, online Meditation, Engagement and Attention values ranging from 0 to 100 were either exported or computed as in Xu et al. (2018). The heatmaps of the eye tracking data along with the screen capture video was recorded using the Tobii Ghost application and Open Broadcaster Software (OBS). Overall objective of this pilot was to collect the multiple sensor data with various spatial and temporal resolutions as well as to synchronize it with the learning logs. This process also aided in determining the data collection and task design complexities. Currently, the data upload module is being used to upload the CSV file of the physiological sensor data of the learner and integrate it with their learning logs.

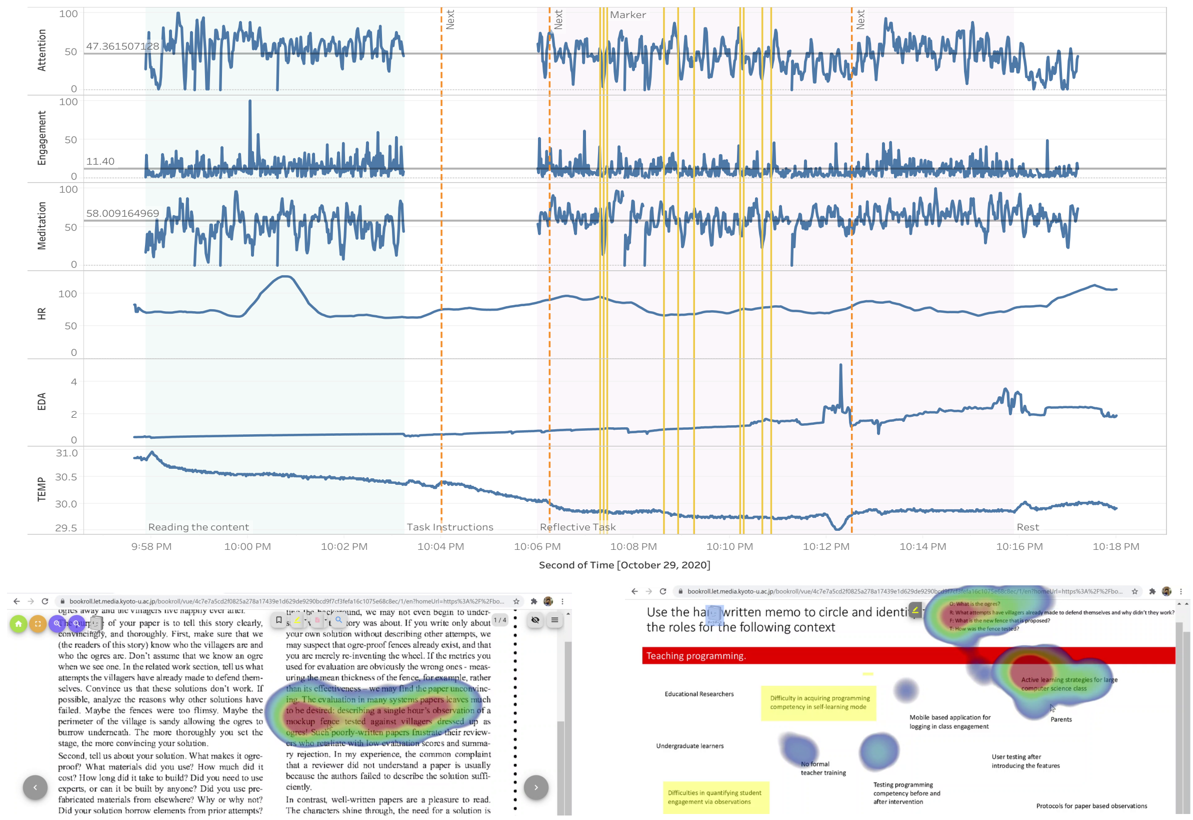


Figure 2: a. Visualization of synchronized multi-modal data captured in GOAL
b. Heatmap of eye gaze during reading and the first task

The future work aims to determine the design requirements for the synchronized data visualization and the pipeline for the data preprocessing in order to enhance the integrated learning analytics services within the GOAL platform. Such a platform would enable further investigation of bio markers for the learning episodes and better utilization of the data streams together to design services to enhance learning.

ACKNOWLEDGEMENTS & REFERENCES

This research was supported by SPIRITS2020 of Kyoto University and JSPS grant 20K20131, 16H06304, 20H01722.

Cukurova, M., Giannakos, M., & Martinez-Maldonado, R. (2020). The promise and challenges of multimodal learning analytics. *British Journal of Educational Technology*, *51*(5), 1441–1449.

Majumdar, R., Yang, Y. Y., Li, H., Akçapinar, G., Flanagan, B., & Ogata, H. (2018). GOAL: Supporting Learner's Development of Self-Direction Skills using Health and Learning Data. In *26th Int. Conf. on Computers in Education Main Conference Proceedings*. (pp. 406-415)

Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., & Yamada, M. (2015). E-Book-based learning analytics in university education. *In Proc. of 23rd Int. Conf. on Computers in Education* (pp.401-406).

Xu, J., & Zhong, B. (2018). Review on portable EEG technology in educational research. *Computers in Human Behavior*, *81*, 340–349. https://doi.org/10.1016/j.chb.2017.12.037

Using IDE-based Learning Analytics to Study Perseverance/Punctuality in Intro Programming

M. Dorodchi, M. Fallahian, A. Benedict, A. Benedict and E. Al-Hossami
(mohsen.dorodchi, mfallahi, abenedi4, abenedi3, ealhossa)@uncc.edu

ABSTRACT: In this paper, we aim to analyze how students learn programming in an introductory course using data from an online Integrated Development Environment (IDE). The goal is to utilize IDE data and learning analytics techniques to provide practical insights and suggestions for course instructors to guide students to success. The data collected from the IDE includes errors tracked in real-time from the time the assignment starts until it is due and even after that once the due date is passed. For each student, the type of error and its frequency are analyzed to understand the behavior of a student during the investigated time period. Then these indicators are compared with other students and the average of the class to check whether it is a student's weakness, or the error is common and arises from the difficulty of the problem. Our analytics highlight the trends between grades and IDE punctuality in such that students who earned ‘A’ and ‘B’ grades are more likely to do and submit their tasks on or before the deadline. On the other hand, students who got lower grades start their tasks on the deadline.

Keywords: Learning Analytics, Topic Modeling, At-Risk of Failing.

1 METHODOLOGY

Analysis of students’ coding patterns in a custom-designed IDE (Dorodchi, 2020) shows how students perform their coding activities and assignments, and how it relates to their learning of programming. It is accessible fully online and integrated with the Canvas Learning Management System (LMS) and an auto grader tool (Codepost.io). The activities include lab activities that were previously completed in closed lab sessions based on an active learning class model (Dorodchi, 2018). The IDE was designed and implemented to help out with the fully asynchronous online course; all lecture and lab activities needed to be done remotely in that environment. Data was then collected within the IDE regarding students’ programming patterns of behavior.

In this research, students’ programming/compiling attempts were collected from students' interactions with the IDE enrolled in an online introductory programming course with 41 students and 44 total activities and assignments. Total of 12,176 compile attempts out of which 63% were unsuccessful as summarized in table 1 were recorded. All the error messages were categorized in four types as shown in table 1 and further passed to a Latent Dirichlet Allocation (LDA) algorithm (Blei, 2003) for topic modeling. Furthermore, these compilation data are used as indicators of punctuality (how the pattern is with respect to due date) and perseverance (how the attempt pattern varies over the assignment period and the entire course). These indicators include the basic static compilation behavior of students versus more complex dynamic situations such as “number of attempts”, “attempt pattern”, and “compilation distribution”.

The “number of attempts” feature was further normalized for each student by the average number of all attempts made by all of the students. We use this relative quantity because using absolute attempt frequency could be misleading because activities/assignments have different difficulty levels. Plotting the attempt of each student from the day at which the task is assigned to the deadline and comparing the attempts of each student with the average number of attempts in a specific assignment reveals four different segments as shown in figure 1(b). Students demonstrated various attempt patterns. The number of attempts, daily attempts, and the number of errors on a particular day is studied to check whether these patterns have any correlation to students’ grades or not. Moreover, this indicator can help educators to give an appropriate amount of time to students to complete their assignments.

Table 1: Some example errors extracted and classified by topic modeling.

Error Type	Sample Error Messages for Different Error Types
Missing Elements	`';' expected, ')' expected, missing return statement`
Incorrect Statements	`reached end of file while parsing, orphaned case`
Variable Problems	`incompatible types, variable choice is already defined`
Runtime Error	`Main method not found in class, ArithmeticException: / by zero`

2 RESULTS AND DISCUSSION

Figure 1(a) shows the attempt indicator for all of the students for all assignments in which value of 1 means the number of attempts made by a student is equal to the average number of attempts in that specific assignment. Consequently, the value of zero indicates no attempt. Figure 1(a) reveals that the attempt indicator of the students with final grade of ‘A’ is greater than one which is more attempts than the average of the class. Using a one-tailed one-sample t-test, we found p = 0.008807 to be significant. This is also true for students with ‘B’, however, the result of one-tailed one-sample t-test results to a higher p-value of 0.04462 which is still significant. Almost no student with a grade lower than ‘C’ has an attempt indicator higher than one. This result highlights that a lower final grade in the course is associated with a lack of practice and experience in programming. Students with these patterns may not provide themselves with the time needed to study and practice, which is particularly important in an introductory programming course. On the other hand, figure 1(b) compares each student’s compilation pattern against the mean of the classroom with respect to the due date. Positive values on the x axis indicates early, zero on-time, and negative late submissions. Figure 1(b) shows at-risk students in the lower-left section of the chart with late submissions and lower number of attempts than average. Students who submit their works on time and attempt more than the classroom’s average to develop successful code are in the upper right section with final grades of ‘A’ and ‘B’. It is worth noting that all the shown grades are without any late penalties.

Figures 2(a) and 2(b) show the compilation distributions for grades ‘A’ and ‘D’, so it enables us to track the compilation behavior of each category from the time the assignment starts until it is due and even after that once the due date is passed. The lines highlight the number of each error category and all attempts during the mentioned time. The figure for those who received ‘A’ highlights that the majority of attempts and errors are made on or before the due date. Evaluation of grades ‘D’ highlights that these students practice and submit their assignment on or after the due date.

Our next step in addition to implementing all the IDE modules is to use a data-driven recommender system through a conversational agent helping out students in programming.

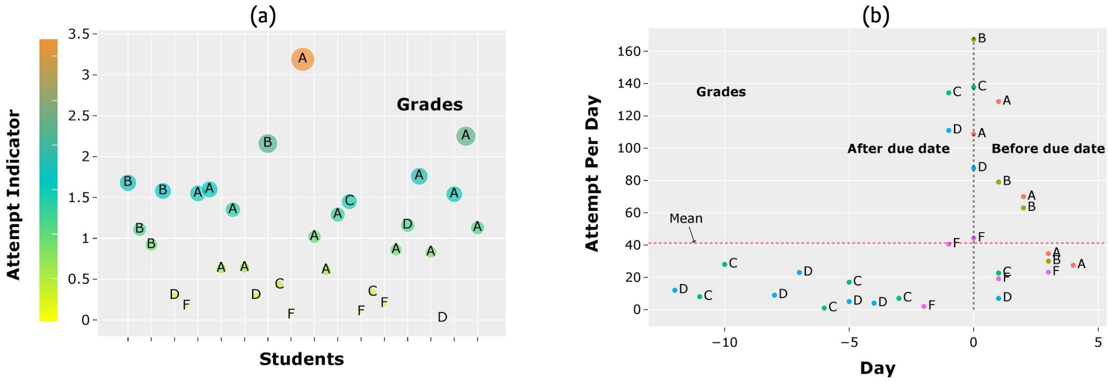


Figure 1: Scatter plot of student grades and a) Attempt per Student; b) Attempt per Day.

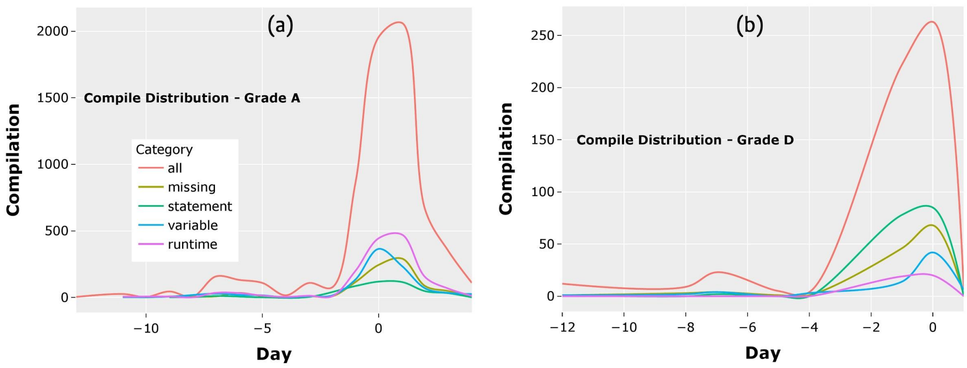


Figure 2: Compilation Distribution for students with grades a) 'A' and b) 'D'

REFERENCES

M. Dorodchi, M. Fallahian, E. Al-Hossami, A. Benedict, A. Benedict. (2020). Custom "Caring IDE" for online offering of CS1. Sixth SPLICE Workshop at L@S 2020.

M. Dorodchi, et al. (2018). Design and Implementation of an Activity-Based Introductory Computer Science Course (CS1) with Periodic Reflections Validated by Learning Analytics, IEEE FIE 2018.

C. D. Hundhausen, D. M. Olivares, A. S. Carter. (2017). Ide-based learning analytics for computing education: a process model, critical review, and research agenda. ACM TOCE 17 (3),1–26.

A. S. Carter, C. D. Hundhausen. (2017). Using programming process data to detect differences in students' patterns of programming. ACM SIGCSE 2017.

M. C. Jadud. (2006). Methods and tools for exploring novice compilation behaviour. ICER'06.

D. M. Blei, A. Y. Ng, and M. I. Jordan. (2003). Latent Dirichlet Allocation. J. Mach. Learn. Res. 3.

Is the Answer Results Always More Informative Than the First Attempt? : An Analysis of Hybrid Data Collected From a Real Sudoku Learning Case

Ci Zhang
Beijing Huo Hua Si Wei Ed Tech LLC, China
zhangci34@outlook.com

Junchen Feng
Beijing Huo Hua Si Wei Ed Tech LLC, China
fengjunchen@huohua.cn

ABSTRACT: Students' performance on the first step might have a more substantial diagnostic and predictive power than their performance on the final result, especially when there is the possibility of teacher intervention. This paper collected formative data from real-time interactive live classes and summative data from homework and assessment on the same learning objective, the Sudoku puzzle. The authors analyzed the students into four groups based on their first step correctness and the final result in the class exercise to show the power of process data. By comparing these four groups of students' subsequent performance, we conclude that the first attempt is sometimes more important than the answer results.

Keywords: Specific learning processes, Hybrid data, Procedural diagnosis

1 INTRODUCTION

Online education is becoming popular, but little literature has analyzed the procedural data during the answering process for diagnoses and evaluations. In China, an online education system combines real-time interactive live courses with corresponding after-class practices. In a live classroom, students learn through teaching activities based on interactive course-ware, including doing exercises. In this way, their operations are recorded, forming big data.

This paper chose the topic "Sudoku" in a math course for children aged 6 to 7 as learning content. Procedural data from live class and consequential data from homework and assessment on the same topic are collected. Researchers found that students' first step data can sometimes be more informative than the actual correctness. Three-parameter logistic model in Item Response Theory (IRT) was applied to binary response data to estimate students' ability (Birnbaum, 1968).

2 DATA COLLECTION

For some problems, if students understand the critical step, it is easy for them to solve it; that is why we chose the Sudoku. Meanwhile, each step of the students' operation directly reflects their thinking on such items. Educators can learn about students' actual thinking from the process data, rather than just guessing based on the result. The dataset consists of the step-by-step actions of 5,362 students on one in-class exercise and the dichotomous responses of those same students on four homework items and an assessment item. All the questions are six by six Sudoku problems, and all students answer those items in a similar pattern. Figure 1 shows the content of the in-class item.

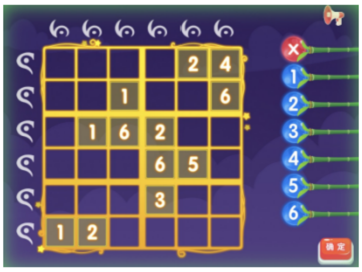


Figure 1: Screen-shot of the in-class exercise

3 ANALYSIS

3.1 Grouping from the in-class exercise

In this lesson, the learning objective is to use a "row-column-block elimination" strategy. The first attempt can somehow reflect students' mastery of this strategy. Therefore, whether the students' first operation meets the teaching requirements is regarded as a classification standard. According to the teacher's instruction, there are six optimal first steps on the in-class exercise. If students' first step is among the six optimal steps, they would be placed in the "first attempt correct" group. Besides, another criterion for grouping is the absolute answer correctness. Under the above two classification standards, students are divided into four groups, as shown in Figure 2.

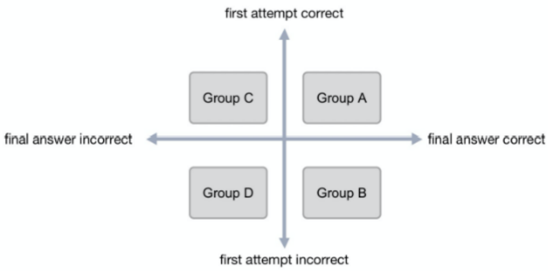


Figure 2: Grouping of students

In general, one would expect the final results to be more informative on assessing students' mastery. However, the process, especially the first step, proves to be more informative. To show that, we examine the impact of the two classification methods through the four groups of students' performance in the subsequent homework and assessment.

3.2 Analysis combined with after-class data

Table 1 displays the correct rate of each item in each group on average. Correctness is defined as finish the Sudoku puzzle in their final submission. By comparing, it is clear that when the final results are the same, students who got the first step correct have a higher correct rate on all items. Researchers generally expect that group B's correct rate should be larger than that of C because group B got the in-class exercise right and group C didn't. However, contrary to the expectations, group C outperformed group B on all items.

To reconfirm, the researchers then applied the 3-parameter logistic model in item response theory on this data, using an R package "TAM" (Robitzsch, Kiefer, & Wu, 2020). The model's expression is: $P(\theta) = c + (1-c)(1+\exp(-Da(\theta-b)))^{-1}$, where P is the probability of correct response for the students; a, b and c are the discrimination, difficulty, and guessing parameters for items, respectively; theta stands for the ability of the students; D is 1.702. Here, the authors estimate the students' ability and calculate each group's mean ability. We also verified that the AUC of using homework results to predict assessment results is 0.886, which means the model's estimations are reliable. Group C's mean ability is significantly larger than group B's (-0.14 vs. -0.21, p-value < 0.01), consistent with the conclusion above.

Table 1: The correct rate for each group on average.

Groups	Description	Home work 1	Home work 2	Home work 3	Home work 4	Assessment
A (*n* = 3,041)	First right, Final right	78.56%	83.69%	78.48%	82.21%	75.87%
B (*n* = 402)	First wrong, Final right	63.64%	73.60%	63.79%	68.62%	63.53%
C (*n* = 1,503)	First right, Final wrong	75.08%	80.04%	75.58%	76.47%	73.02%
D (*n* = 416)	First wrong, Final wrong	62.95%	70.57%	65.49%	67.33%	59.18%

4 DISCUSSION

Why did the children in Group B perform less well on after-class practices than Group C? The authors guessed that it was the teacher's intervention. Generally, teachers may help with less able students but rarely intervene in the first step. It is possible that the students in Group B only answered the classroom exercises correctly with the teacher's help but did not fully grasp the teaching objectives. So it makes more sense to diagnose the first step.

To sum up, we believe that students' in-class performance on the first attempt might be more informative than the absolute correctness, mainly when intervention exists. This conclusion may also apply to other problem solutions where key strategies are repeatedly used. In the past, diagnosis and analysis of knowledge mastery mostly used consequential data. This paper hopes to remind researchers that using only the resulting data is somewhat risky when it is unknown if students completed the exercises independently.

REFERENCES

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 395-479). MA: Adison-Wesley.

Robitzsch, A., Kiefer, T., Wu, M. (2020). *TAM: Test Analysis Modules.* R package version 3.5-19, https://CRAN.R-project.org/package=TAM.

Blockchain-based Data Verification and Consent Management for Trusted Learning Analytics Using Mentoring Chatbots

Peter de Lange, Lennart Bengtson, Alexander Tobias Neumann, Ralf Klamma
RWTH Aachen University, Aachen, Germany
{lastname}@dbis.rwth-aachen.de

ABSTRACT: Automating mentoring by intelligent chatbots to give students more support in self-regulated learning in higher education is a promising idea. However, to make the mentoring bots intelligent they need access to sensitive personal learning analytics data, possibly decreasing the trust of students in the mentoring processes. We have created and evaluated two important contributions to increase the trust. First, a blockchain-based verification process for increasing the transparency of data access and second, a consent self-management of learning analytics data, retaining the control in the hands of the students. Evaluation results are very positive and will lead to the application in a large-scale mentoring support environment.

Keywords: Blockchain, trusted learning analytics, consent management, chatbots.

1 INTRODUCTION

Despite being a vital component of educational processes, mentoring relationships in today's higher education are hard to maintain. This is caused by the mismatch of the number of mentors in relation to the number of mentees enrolled in higher education programs. Consequently, this creates the need for scalable technical support for automated mentoring processes (Klamma et al., 2020). In addition to the right tools, learning analytics (LA) that require students' personal data, make mentoring effective. This LA data is sensitive, as it reflects the learning behavior, which involves personal information such as grades. Consequently, the consent of each individual learner has to be requested prior to the extraction and analysis of their personal data. Security and transparency measures are required for collection and processing of the data, to fulfill data processing regulations such as the European GDPR. Ideally, this also increases the learners' trust in and acceptance of automated mentoring by keeping control over their data in their hands, confidential and secure. In this contribution, we tackle the issue of enforcing the students' preferences on data processing by using blockchain technology. Thereby, we maintain a record of the learners' accessed personal data and record their consent to the collection and processing of their data. Our approach utilizes chatbots as conversational interfaces to manage consent and access personal LA data. It is driven by the following two research questions:

RQ1: Can consent self-management contribute to an increase of perceived control over data?

RQ2: How does verification of the integrity and source of LA data benefit trust in the processing and analysis of personal data?

2 BACKGROUND AND RELATED WORK

Data stored on a public blockchain is practically immutable, yet still openly accessible. Using these characteristics, various approaches for storage and verification of documents have been suggested. The process is similar in most cases: a document of any kind is issued, hashed, and the hash is stored on a blockchain. When a third party wants to verify the integrity of a received document, this document is hashed and through comparison of the two hashes, it can be verified.

This approach is particularly interesting in education, as the process of verifying the authenticity of conventional degrees in physical form can be slow and tiresome (Jirgensons and Kapenieks, 2018). Thus, the benefit of this form of certificate management is the simplicity of automatic and independent verification, compared to traditional physical degrees. Therefore, various proposals were made for applying blockchain to certificates in education. Exemplary, we want to mention Blockcerts, as one of the first major contributions in this field. The specific combination of LA and blockchain technology has not received much attention in research yet. (Forment et al., 2018) express their intent to explore ideas on how blockchain can be used to deal with privacy challenges and concerns. According to the authors, smart contracts should be used to “manage the access to the information” in accordance with agreements concerning data stewardship and ownership, between the stakeholders involved. However, in another contribution, they express concerns related to privacy as “in Blockchain all data is public” (Amo et al., 2019) and all users in the network would have access to all transactions recorded about the access to a learner's data.

3 CONCEPTUAL APPROACH AND IMPLEMENTATION

On a conceptual level, our contribution requires the implementation of three main parts, the verification of LA data, the consent management and the chatbot as conversational interface for learners to access their LA data logs and manage their consent. To achieve the verification, we use a blockchain-based registry that stores references to all LA data extracted from the Learning Management System (LMS). This process of storing a reference and comparing against that reference later is a widely established use-case for blockchain technology. By storing hashes for reference only and relying on the irreversible nature of hash-functions, we discourage access to personal data on the blockchain. For consent management, we provide learners with an interface that allows direct interaction with the system. Issued consent is stored on the blockchain and is integrated into the extraction process to restrict access to personal data, for which no consent was given. Our concept includes a chatbot for learners to use the features for consent management and verification of LA data. This bot enables the storage and revocation of learners' consent, the display of their currently supplied consent and a listing of their collected personal data.

Fig. 1 gives an overview of our implementation by depicting an exemplary usage scenario. At the beginning of the course, Alice and Bob are both informed that their consent is required before they can use the personalized LA features. They now store their consent to the collection of personal learning data from the LMS with the help of the chatbot interface and then proceed to use the LMS course room. Before data is extracted, Alice's and Bob's consent is queried from the consent registry, and only the data the consent was given for is then stored in a Learning Record Store (LRS). Alice and Bob both can ask the bot to display their collected personal data from the LRS at any time. The bot then provides this data. Based on a comparison with the reference in the verification registry, it is indicated whether the data can be traced back to an extraction from a source respecting the learner's consent (*verified data*) or not (*unverified data*).

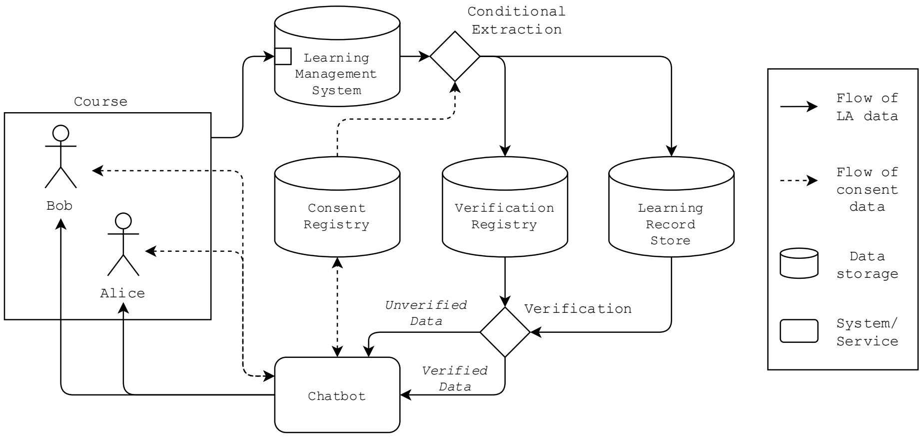


Figure 1: Usage scenario for verification and consent management.

4 PRELIMINARY EVALUATION AND OUTLOOK

We evaluated our approach with both mentees (learners) and mentors (teachers). Altogether, we recruited 10 students and 6 faculty members of multiple universities, and we thus conducted 16 online evaluation sessions. The participants went through a complete process similar to the one sketched in our usage scenario in Fig. 1. This preliminary evaluation showed the applicability of our approach and both mentees and mentors valued the idea of having verifiable consent-management and LA data processing. The chatbot provided low barrier access to both features (AVG=4.19, SD=1.13). Even if only applied in a simulated mentoring environment, with an average of 4.63 (SD=0.59), participants found the consent self-management functionalities to increase their perceived control of what LA data was recorded (RQ1) and, with an average of 4.19 (SD=0.88), stated that the trust towards the secure handling of their LA data was increased due to the verification features (RQ2). We are confident that this approach bears the potential of providing a step towards explainable consent self-management, while also raising the acceptance and trust of students towards learning data processing, leading to more trusted learning analytics. We are currently evaluating our approach within the scope of a large German research project and will report back on this once we have first results in future work.

Acknowledgments The authors would like to thank the German Federal Ministry of Education and Research for their kind support within the project ``tech4comp" under the project id 16DHB2110.

REFERENCES

Amo, D.; Fonseca, D. F., Alier, M., García-Peñalvo, F. J., Casañ, M. J., & Alsina, M. (2019). Personal Data Broker: A Solution to Assure Data Privacy in EdTech. In: LCT'19, Vol. 11590, LCNS, 3–14.

Forment, M. A., Filvà, D. A., García-Peñalvo, F. J., Escudero, D. F., & Casañ, M. J. (2018). Learning Analytics' Privacy on the Blockchain. In: TEEM'18, ACM, 294–298.

Jirgensons, M., & Kapenieks, J. (2018). Blockchain and the Future of Digital Learning Credential Assessment and Management. In: J. of Teacher Education for Sustainability 20 (1), S. 145–156.

Klamma, R., de Lange, P., Neumann, A. T., Hensen, B., Kravcik, M., Wang, X., & Kuzilek, J. (2020). Scaling Mentoring Support with Distributed Artificial Intelligence. In: ITS'20, Vol. 12149, LCNS, 38–44.

Do student advisors prefer explanations using local linear approximations (LIME) or rules (LORE) in the prediction of student success?

Tinne De Laet & Lotte Huysmans
Leuven Engineering and Science Education Center (LESEC), KU Leuven, Belgium
first.lastname@kuleuven.be

ABSTRACT: Student advisors want to provide aspiring students with appropriate recommendations or remedial actions in their transition from secondary to higher education. A prediction of student success might provide support in this process. However, as student success prediction models are often black box models, there is a need for explainable AI that helps student advisors to understand why a student has a higher chance of failure or success. Understanding why a black box model makes a certain prediction is also important in assessing trust, which is fundamental when actions will be taken based on the model's predictions. This work makes a comparison between different explainers supporting a black box prediction of student success. The comparison focuses on two model-agnostic explainers: Local Interpretable Model-agnostic Explanations (LIME) and LOcal Rule-based Explanations (LORE). The poster will present an evaluation with end users (student advisors). The results indicate that advisors prefer working with a LIME explanation over working with a LORE explanation because of the extra nuanced information in the LIME explanation.

Keywords: explainable AI, student advising, student success, prediction

INTRODUCTION

When transitioning from secondary to higher education students have a large number of options to choose from. Good guidance of these aspiring students when making their study choice and transition towards university is thus important. Many universities use professionally trained advisors to provide this guidance to students. They can help aspiring students with study advice and often supplement this task with support of enrolled students during their study career, or even with content-wise tutoring on first-year subjects. Learning Analytics (LA), defined as "the measurement, collection, analysis, and reporting of data about the learners and their contexts for the purposes of understanding and optimizing learning and the environment in which it occurs" (Long 2011) also has the potential to support the advising of aspiring students and to support advisors in the process of the advising. LA can be used to predict the probability of a student to succeed in their first year at university. These predictions can then be used by the student advisors to help the conversations they have with students. Some prediction models, such as single decision trees or linear models, are transparent and easy to interpret by a user. The problem with these models is that they often have a low predictive power, meaning that it is difficult to accurately predict the success of a student. Other models are less transparent but have more predictive power. As the models that predict a student's probability of success are used to advice students it is not sufficient to only know whether a student has a high likelihood of success or failure but it is important to understand why this is the case. Therefore, there is a need for some kind of explanation of the decision made by the complex predictive model. Explainable AI (Adida 2018) provides methods and techniques such that results of AI models can be understood by end-users. These methods or techniques can be applied after the prediction model to explain how the model came to its decision, hereby having the potential to make black-box predictions easier to understand thus providing an additional tool for student advisors when advising aspiring students.

METHODOLOGY

The data used is obtained from the students in a first bachelor of Engineering Science from the year 2015 till 2019 (four consecutive academic years) and consists of prior academic achievement (in math, physics, and chemistry), the number of hours of mathematics education and the self-reported effort level in secondary school, learning and studying skills (goal and affective strategies) measured using a validated questionnaire, and as an outcome measure, the study results after the first semester. An XGBoost classifiers was trained and two explainers were used to explain the results of the black-box prediction: Local Interpretable Model-Agnostic Explanations (LIME) (Ribeiro 2016) and LOcal Rule-based Explanations (LORE) (Guidotti 2018). Both explainers are model-agnostic approaches. A LIME explanation provides insights in how much every feature contributes to the classification of the student. A LORE explanation provides a decision rule and some counterfactuals, where the counterfactuals indicate how much a feature must change in order for the student to be classified in the opposite class.

		LORE			
Features		Decision rule		counterfactual	counterfactual
Mathematics grades	>90%	X	>8.5	<60%	
Physics grades	>90%				
Chemistry grades	80-90%	X			<60%
Hours of math	8h				
Effort level	high	X			
Affective strategies	high-medium				
Goal strategies	low-medium				

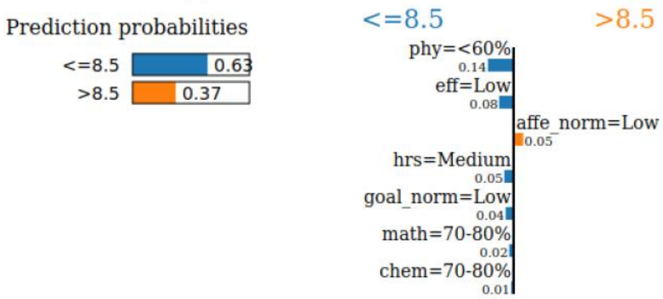


Figure 1: Example of a LORE (top) and LIME (bottom) explanation where a student is predicted as not at risk.

In a user study we assess the preference of fourteen end-users towards LIME or LORE explanations. To this end a survey was combined with interviews. The survey itself consists of a LIME part and a LORE part, each containing question blocks for three students. The students are selected based on a correct risk class prediction by the black box model. For every student the prior academic knowledge and soft skills are provided to the student advisors. First they were asked to assess whether they think if that student would fall in the at-risk or not-a-risk class and which of the students characteristics contributes to the risk classification. This enforces the student advisors to create an opinion about that student before they have seen a prediction or explanation. Then they are provided with two LIME or two LORE explanations and asked to select the explanation that resembles the most their earlier assessment. All advisors were asked at the end if they had a preference for either LIME or LORE and to elaborate on the reasons for their preference. Some of the student advisors participated in an interview after the survey which consisted out of two questions where they are provided with an explanation and are asked how they would use that explanation to advice the aspiring student.

RESULTS

From the fourteen participants only three preferred to work with the LORE explanations and their main reason was that it is easier to interpret than LIME explanations. One of the student advisors who preferred LORE explanations stated the following: "From a LORE explanation it is immediately clear what changes lead to a change of class, while to deduce the same information from a LIME explanation you have to start calculating". Another student advisor that preferred LORE over LIME stated: "Although I'm more familiar with LIME, I think it is easier to conceptually depict the reality of the student from a LORE explanation." The student advisors that prefer working with LIME explanations pointed out two main reasons. One is that LIME provides more nuanced information about the student compared to a LORE explanation, as beautifully stated by one of the student advisors: "The LIME representation is more nuanced and easier to interpret as it shows for every feature how the feature contributes to the classification". The second reason is that in a LIME explanation information about every feature is present, where a LORE explanation only provides information on the features that are present in the decision rule. One particular aspect that is used a lot by the student advisors when looking at a LIME explanation is what feature contributes to which class, information that is not readily available in a LORE explanation. Of this small test group the majority thus preferred LIME over LORE to prepare a conversation with a student and thus preferred probabilities over causes. Although some of the student advisors who preferred LIME did also mention that if they have to show one of the two explanations to a student that they would prefer to show the LORE explanation as it is the easier visualization to explain to the student.

CONCLUSION

The poster will present the result of a case study of explainable LA with student advisors where two explainable AI methods, LORE AND LIME, were compared. The results indicated that advisors prefer working with a LIME explanation over working with a LORE explanation for a student because of the extra nuanced information provided in the LIME explanation.

REFERENCES

Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). IEEE Access, 6, 52138–52160.

Guidotti, R., & Monreal, A. Local rule-based explanations of black box decision systems, https://ui.adsabs.harvard.edu/abs/2018arXiv180510820G, 2018

Long, P., & Siemens, G. Penetrating the fog; analtyics in learning and edcuation. Educause Review, 2011.

Ribeiro, M. T., & Singh, S. "Why should i trust you?" explaining the predictions of any classifier. *KDD p.1135-1144, ACM*, 2016.

Exploring Model Performance for Racial and Gender Groups in Predictions of Algebra Outcomes in Middle School

Mia Almeda, Megan Silander, Joshua Cox
Education Development Center
malmeda@edc.org, MSilander@edc.org, JCox@edc.org

ABSTRACT: In response to emerging research that has identified biases in machine learning models (Paquette et al., 2020), there is an increasing interest within the learning analytics community to use demographic information to examine issues of equity and identify the limitations of these models. In this study, we developed a prediction model of students' 8th algebra outcomes based on 6th grade attributes related to math assessment scores, student demographics and school characteristics. We evaluated how well our model generalizes to subgroups of race and gender. Our prediction model performed generally well across different demographic groups, but with slightly less accuracy for Black and Hispanic students. These findings have implications for supporting students' understanding of algebra and conducting future work that tests and validates models of math learning for varied demographic groups in middle school.

Keywords: Prediction modeling, algebra learning, demographic cross-validation

1 INTRODUCTION

There is burgeoning interest within the learning analytics community to examine issues of equity as machine learning models have been shown to reproduce racial prejudices (Chouldechova, 2017) and gender biases (Bolukbasi et al., 2016). Paquette and colleagues (2020) argue that there is scarce research using demographics to assess model bias through testing and validation in order to investigate whether machine learning models perform worse for certain groups of students. The present study addresses this research gap by investigating how a prediction model of algebraic achievement in 8th grade generalizes to different demographic groups in middle school. As part of the *Math Data Collaborative* project funded by Schmidt Futures, we leverage NWEA MAP Growth interim mathematics assessments to predict algebra learning in middle school and use subgroups of race and gender as different testing sets to evaluate model accuracy. School districts typically use interim assessments to measure student growth and project proficiency on state accountability tests and to identify schools in need of remediation, and teachers also use results to target instruction to improve student learning. It is critical to evaluate whether model accuracy differs for different demographic subgroups of students as findings from these models could influence how teachers and schools identify which students are at-risk of failure and in need of more resources. For example, biases in the prediction models using these assessments could lead to a greater likelihood of assigning students in these subgroups to the wrong intervention groups.

2 SAMPLE AND MEASURES

Our full sample consists of 332,147 middle school students in 1490 schools who have 6th grade MAP Growth mathematics achievement test scores in 2013-2017. To ensure that the same math outcomes are being assessed, the analytic sample is limited to students in Common Core-aligned states. Our analytic sample includes 150,547 students in 810 schools within a midwestern Common Core-aligned state who have both 6th and 8th grade mathematics assessment scores. MAP Growth data include student scores on 6th grade fall term assessments, broken down by four skill areas: 1) Real & Complex Numbers, 2) Operations & Algebraic Thinking, 3) Geometry, and 4) Statistics and Probability. 8th grade winter term scores for Operations & Algebraic Thinking are used as the outcome in our model. Focal math skills scores were standardized within the analytical sample (z-scored) and aggregated per school and cohort (average z-scores per school).

Student-level demographic measures include race, gender, age when entering 6th grade, whether student changed schools between 6th and 8th grade, instructional week when the assessment was taken, whether student repeated, and cohort (AY 2013-2017). School-level demographic measures include school urbanicity, type of school, school size, percentage of racial/ethnic enrollment, and percentage of students eligible for free or reduced lunch.

3 METHODOLOGY

We built a model of 8th grade algebra outcomes based on 6th grade predictors (assessment scores, student- and school-level demographics) using RapidMiner Studio 9.3 data mining software using the Weka M5P algorithm. W-M5P is a reconstruction of the M5 algorithm that induces a decision tree and incorporates linear regression models (LMs) at the leaves. The model was developed using 4-fold, student-level batch cross-validation. This approach repeatedly trains the model on 75% of students and tests on the remaining 25% to estimate model generalizability to new data. The model was evaluated using cross-validated correlations (Pearson's *r*) between the model and the data. Positive cross-validated correlation indicates that the relationship is consistent between the training and test datasets but has no implication about the relationship's direction.

After building and cross-validating the model, we took the model's prediction on the test sets across every fold and assessed model performance on subsets of the data based on student demographic characteristics. In particular, we compared performance of W-M5P model by race (Black, Asian/Hawaiian, Hispanic, American Indian/Other, or Multi-Ethnic versus White) and gender (female versus male).

4 RESULTS

As shown in Figure 1, the W-MP5 regression tree generated three splits and five terminal nodes. Out of the different 6th grade predictors (assessment scores, student- and school-level demographics), the hierarchy of the regression tree indicates that 6th grade scores in Real & Complex Numbers, followed by Operations & Algebraic Thinking, are the strongest predictors of algebra performance in 8th grade.

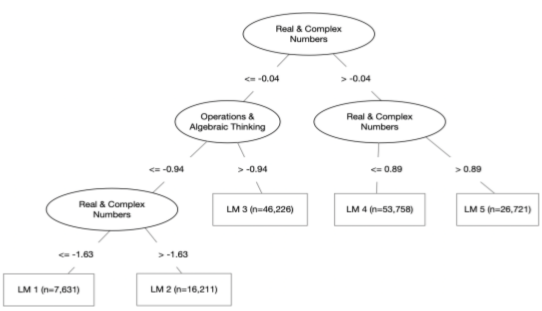


Figure 1. Decision tree generated by the W-M5P algorithm

We tested the cross-validated regression model by student demographic comparisons: race and gender (see Table 1). This model performed marginally worse for Black and Hispanic students as compared to that of White students. However, model performance for Black and Hispanic students were still quite good when examining their cross-validated correlations. In contrast, it was slightly more accurate for Asian/Hawaiian and Multi-Ethnic students compared to White students. Students who identified as American Indian/Other yielded the same model performance as those who identified as White. When examining gender differences, our W-M5P model achieved a slightly lower cross-validated correlation for females relative to males. In general though, the W-M5P model performed relatively well across different demographic groups. Predictions for each subgroup were close to each other and were comparable to the overall model (*r* = 0.833).

Table 1: Performance of W-M5P models of algebra outcomes for demographic groups.

Group	N	Cross-validated Correlation
White	66,496	0.821
Black	21,581	0.796
Asian/Hawaiian	6,975	0.845
Hispanic	37,617	0.799
American Indian/Other	13,233	0.821
Multi-Ethnic	4,645	0.848
Male	74,106	0.838
Female	76,441	0.828

5 DISCUSSION

Our W-M5P model suggests that building students' skills in Real & Complex Number Systems at the beginning of middle school is likely important for developing their understanding of algebra two years later, more so than Operations & Algebraic Thinking. These findings provide initial evidence that instruction focusing on Real & Complex Number Systems, such as number properties, operations & ratios, might be particularly important in getting students ready for success in algebra. Additionally, findings from evaluating the reliability across the different demographics indicate less model accuracy for Black and Hispanic students. Although this pattern of results is not surprising, the slightly lower reliability for demographic groups who have been historically and systematically underserved is concerning. Our findings imply that our model can be used without a high risk of compromising equity in predictions, but it can also be improved on to increase model accuracy for Black and Hispanic student populations.

In future work, we plan to extend our research in demographic cross-validation by incorporating district policy responses to COVID-19 in our model and examining how these relate to differences in math achievement in a national cohort of middle-grades students. As we investigate which district policies support the math learning of student populations disproportionately impacted by the pandemic, exploring the different ways in which demographic variables influence model performance of algebra learning is an important step to identify potential limitations of learning analytics models and to promote their equitable use in middle school education.

REFERENCES

Bolukbasi, T., Chang, K. W., Zou, J., Saligrama, V., & Kalai, A. (2016). Man is to computer programmer as woman is to homemaker? In *Advances in Neural Information Processing Systems*, Barcelona, Spain, 4349-4357.

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, *5*(2), 153-163.

Paquette, L., Ocumpaugh, J., Li, Z., Andres, A., & Baker, R. (2020). Who's Learning? Using Demographics in EDM Research. *Journal of Educational Data Mining*, *12*(3), 1-30.

Visual Attention Patterns on Dashboard during Learning with SQL-Tutor

Faiza Tahir, Antonija Mitrovic, Valerie Sotardi
University of Canterbury
faiza.tahir@pg.canterbury.ac.nz, {tanja.mitrovic, valerie.sotardi}@canterbury.ac.nz

ABSTRACT: In this study, we investigate how students use a dashboard in SQL-Tutor, an intelligent tutoring system that teaches the SQL query language. The dashboard is shown each time the student solves a problem, illustrating the student's progress both in graphical and a text-based form. The analyses of students' eye-tracking data show that students give much attention to the dashboard, especially when the dashboard is shown for the first time. In subsequent situations, students tend to focus on goal progress and the visualization of the student model. These results will help us to refine the dashboard in SQL-Tutor with important visualizations.

Keywords: Learning analytics, dashboard, intelligent tutoring system, eye tracking.

1 INTRODUCTION AND BACKGROUND

Learning analytics includes effective ways of measuring, analyzing and reporting learning outcomes students achieve in various learning environments. Learning analytics dashboards are tools which effectively visualize learning information to students and teachers (Verbert, Duval, Klerkx, Govaerts, & Santos, 2013). The purpose behind presenting this information is to make students aware of their learning progress, increase their self-monitoring and reflecting skills, and regulate learning strategies (Bodily et al., 2018), with the goal of improving learning. Many studies explore effects of visualizations used in dashboards through learner's eye-tracking data analysis. Eye tracking studies mainly focus on differences between novices and experts. For example, Barral et al. (2020) investigate how various types of students process information represented by charts and graphs, and provide evidence that adaptive guidance provided in the form of narrative-based charts visualization can benefit both novices and advanced students. Another eye tracking study compared nine different notational visualization systems, and reported that students preferred simple, easy and straightforward notations (Vatrapu, Reimann, Bull, & Johnson, 2013). Goldberg and Helfman (2011) revealed that students retrieved information easily from linear graphs organized either vertically or horizontally, rather than radial graphs. Another project combined eye-tracking data and students' interactions (log data) to analyze the perceptual speed, visual working memory, spatial memory, and visual scanning of students on different visualizations (Conati, Lallé, Rahman, & Toker, 2020). The authors suggested the importance of gaze data in user modelling and as a predictor of user interactions. All these studies investigated students' perceptions and preferences of different visualizations, and also the potential for predicting future performance. However, there is no research on how frequently dashboards should be shown to students to influence their skills and student interaction with dashboards in my opinion. In this contribution, we provide the initial results of an eye tracking study conducted to fill these gaps. The context of this study is SQL-Tutor, a mature intelligent tutoring system (ITS) for teaching problem solving in SQL (Mitrovic, 2003).

2 STUDY DESIGN AND PROCEDURE

We extended SQL-Tutor to support the three phases of Zimmerman's (1990) theory pf self-regulated learning, which targets goal setting, self-reflection, and monitoring skills of students. To support self-regulation, we added a dashboard to SQL-Tutor. The dashboard is presented to the student upon completion of a problem (Figure 1). The top section of the dashboard provides the overall information about the student's history, such as his/her pre-test score, current knowledge level, total time spent with SQL-Tutor, total problems solved with SQL-Tutor, the highest problem complexity, and the percentage of attempts on which the student required to see the complete solution. The second section of the dashboard visualizes the student's progress and the average class progress on each goal in the form of skill meters. If the student has achieved the current goal, the dashboard shows an appreciation message along with the next goal selection option; otherwise, it shows two strategies to select the next problem. The bottom section of the dashboard presents two graphs, which track the problems completed and time spent with SQL-Tutor per week. The last component is the open student model, i.e. the visualization of the student's knowledge in terms of six clauses of the SQL Select statement (select, from, where, order by, having, and group by).

Figure 1. Dashboard of SQL-Tutor

The participants recruited for the study were undergraduate or postgraduate Computer Science students who had previous experience of problem solving in SQL-Tutor. At the beginning of the session, the participant was asked to sit in front of the Tobii eye tracker and calibration test was completed. After calibration, the participant worked with SQL-Tutor while their gaze data were recorded. The students were not required to solve a specific number of problems, but were required to work for 30 minutes.

3 PRELIMINARY INSIGHTS AND FUTURE WORK

To examine the visualization patterns and student attention on the dashboard, we identified parts of sessions where the dashboard was shown for the first time (phase 1) or last time (phase 3), as well as from the middle of the session (phase 2), when the student has completed five problems. Even though the students are still being recruited and data analysis is not completed, we can already deliver some preliminary insights. Students spent an average of 15s (sd = 16) looking at the dashboard in Phase 1, when it was first presented to them. However, this time declines on subsequent viewings (phases 2 and 3). The gaze patterns revealed that students looked at all three sections of the dashboard in phase 1. However, in phases 2 and 3, they only focused on the current goal progress, learning strategies, and the open learner model, as illustrated in Figure 2. Students looked at the graphical presentation of completed problems instead of text-based in their subsequent viewings, which shows their preferences. An interesting finding is that students did not pay attention to the class progress until they achieved a goal. Once they achieved a goal, they not only spent more time on the dashboard and looked at class progress but they also focused on their open learner model to assess their knowledge. The highest problem complexity measure on the dashboard did not receive attention. The possible explanation of this could be that the study was voluntary and students were not motivated to solve complex problems. These initial findings give us insights into student preferences for various elements of the dashboard and some indications on how frequently they want to see the dashboard. Further analysis of these results will help reveal the reasons for such behaviors and refine the dashboard.

Figure. 2 Aggregate eye gaze pattern after solving five problems

REFERENCES

Bodily, R., Kay, J., Aleven, V., Jivet, I., Davis, D., Xhakaj, F., & Verbert, K. (2018). *Open learner models and learning analytics dashboards: a systematic review.* Proceedings of the 8th International Conference on learning analytics and knowledge, pp. 41-50.

Barral, O., Lallé, S., & Conati, C. (2020). *Understanding the effectiveness of adaptive guidance for narrative visualization: a gaze-based analysis.* Proceedings of the 25th International Conference on Intelligent User Interfaces, pp. 1-9.

Conati, C., Lallé, S., Rahman, M. A., & Toker, D. (2020). Comparing and Combining Interaction Data and Eye-tracking Data for the Real-time Prediction of User Cognitive Abilities in Visualization Tasks. *ACM Transactions on Interactive Intelligent Systems, 10*(2), 1-41.

Goldberg, J., & Helfman, J. (2011). Eye tracking for visualization evaluation: Reading values on linear versus radial graphs. *Information visualization, 10*(3), 182-195.

Mitrovic, A. (2003). An intelligent SQL tutor on the web. *International Journal of Artificial Intelligence in Education, 13*(2-4), 173-197.

Vatrapu, R., Reimann, P., Bull, S., & Johnson, M. (2013). *An eye-tracking study of notational, informational, and emotional aspects of learning analytics representations.* Proceedings of the 3rd International Conference on Learning Analytics and Knowledge, pp. 125-134.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist, 57*(10), 1500-1509.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational psychologist, 25*(1), 3-17.

Predicting medicine students' achievement and analyzing related attributes with ANN and Naïve Bayes

Diego Monteverde-Suárez[1], Patricia González-Flores[1], Roberto Santos-Solórzano[1], Manuel García-Minjares[1], Irma Zavala-Sierra[1], Melchor Sánchez-Mendiola[1]
Universidad Nacional Autónoma de México[1]
dmonteverde@unam.mx, {patricia_gonzalez, roberto_santos, manuel_garcia, irma_zavala, melchor_sanchez}@cuaieed.unam.mx

ABSTRACT: This study tests and compares the use of artificial neural networks (ANN) and Naive Bayes to predict students' academic success at the end of the first year in an undergraduate program in medicine and to identify the attributes relevant to prediction. Both methods attained similar predictive results (greater than 70% precision, sensitivity and sensibility). Naive Bayes confirmed students' prior knowledge as the most important attribute for prediction. Variables related with the upper secondary school were found to have a greater incidence in predicting the students' irregularity, while those regarding their knowledge, in students' regularity. Further research is needed to study if curriculum, teacher expectations, or a bias in the model explain these findings.

Keywords: medical education, Naïve Bayes, educational data mining, artificial neural networks

1 INTRODUCTION AND CONTEXT

Dropout and slow academic progress are persistent problems in a six-year undergraduate program in medicine (Campillo Labrandero et al., 2019). In order to implement interventions to promote academic success, it is necessary to identify at-risk students early on and know their attributes. Using statistical methods, several research projects have identified psychological factors and prior academic achievement as the main variables related to academic performance of students in this program (Gatica-Lara et al., 2010; Urrutia Aguilar et al., 2014), as well as the incidence of the type of upper secondary school program where the students studied (Nieto Domínguez et al., 2003).

Although educational data mining or EDM has been used successfully in several disciplines for predicting academic success (Abu Amra & Maghari, 2017; Devasia et al., 2016; Koedinger et al., 2015; Mhetre & Nagar, 2018; Yukselturk et al., 2014), no such studies were found for medical education. This research tests and compares two different EDM methods to predict students' academic performance in an undergraduate program in medicine and identify attributes relevant to prediction. Artificial neural networks (or ANN) and Naïve Bayes were chosen because they have been reported to achieve good results for these purposes and they represent different approaches.

2 METHOD

This study used data from 7,976 anonymized students from the 2011 to 2017 cohorts, collected upon their enrollment to the program, as well as from students' records. The dataset comprises 47 categorical, discrete numerical and continuous numerical variables regarding students' demographics, family environment, socio-economic status, prior educational trajectory, students' initial knowledge in eight subjects, academic achievement in the first year and student's identifier. A full description of the variables is published in: https://predacademicachv.wixsite.com/results. The dependent variable was students' academic success at the end of the first year and two situations were defined: (i) regularity: students who complete all the required courses for the first year (value 1), (ii) irregularity: those who fail one or more of these courses (value 0).

We started by eliminating 910 records with missing data and generating two databases. We pre-processed each database according to the requirements of the specific method (i.e. converting categorical to numerical values for Naïve Bayes and replacing missing values with smooth imputations for ANN). Then, we divided the two datasets into a "training set" consisting of 80% randomly selected student records, and a "test set" with the remaining 20%.

With each of the methods, we estimated two models, one to predict students' regularity and the other, their irregularity. In the case of ANN, a Multilayer Perceptron (MLP) neural network with backpropagation (BP) with two hidden layers was used. For Naïve Bayes, using R, we calculated the probability and the score for each category of the variables, the score for each student, and the *epsilon* values to identify the relevance of the attributes for prediction:

$$\epsilon(X_i) = \frac{N_{X_i}[P(C_k|X_i) - P(C_k)]}{[N_{X_i}P(C_k)(1-P(C_k))]^{\frac{1}{2}}}.$$

Where C_k represents the class, X_i the attribute in accordance with the response category and N_{X_i} the number of students with attribute X_i. We determined the threshold by plotting the sensitivity and precision, and selecting the point in which both values were closest to 1. Furthermore, to define the profiles of regular and irregular students, we carried out an analysis of the sensitivity of the different variables in the ANN models and studied the estimated parameters $\epsilon(X_i)$ from the Naïve Bayes models.

3 RESULTS AND DISCUSSION

Both models' predictive results were similar and equal to or greater than 70% in sensitivity, specificity and precision, thus supporting previous studies on ANN and Naïve Bayes.

Table 1. Comparison of the models' results for both classes

Focus	Model	Sensitivity	Specificity	Precision
Irregularity	Neural networks (ANN)	0.72	0.75	0.74
	Naïve Bayes	0.72	0.70	0.71
Regularity	Neural networks (ANN)	0.75	0.72	0.73
	Naïve Bayes	0.70	0.72	0.71

In regards to contributing factors to prediction, while it was not feasible to define satisfactory student profiles with the ANN sensitivity analysis, Naïve Bayes made it possible to compare the predictive value of the variables for regularity and irregularity. In both cases, prior knowledge was the attribute with greater predictive value. When comparing the order of variables based on *epsilon*, 22 of them play an identical or similar role in the prediction, but two attributes (upper secondary type and campus) have a greater relevance in predicting the students' irregularity, and five variables regarding prior knowledge in mathematics, literature, grammar, reading comprehension and vocabulary are more important for the prediction of students' regularity. An interactive version of the results is published in: https://predacademicachv.wixsite.com/results.

These findings give way to various observations. The relevance of a student's prior knowledge in specific disciplines is corroborated as a predictor in students' academic success in medicine and stresses the urgency of implementing interventions to strengthen this knowledge. Secondly, these results raise the question of why the upper secondary type is a good predictor of irregularity but not for regularity, and point out the need and direction of further research. Initial hypotheses could be that the curriculum creates a disadvantage for students in specific types of upper secondary, or that teachers' expectations on the success of students from each type of program have an impact on their performance ("Pygmalion effect"). Likewise, a bias in the Naïve Bayes models towards a specific upper secondary type could explain the differential behavior of the prediction, which would mean it is important to study whether there is some type of discrimination in the algorithms.

REFERENCES

Abu Amra, I. A., & Maghari, A. Y. A. (2017). Students Performance Prediction Using KNN and Naïve Bayesian. In IEEE (Ed.), *8th International Conference on Information Technology (ICIT)* (pp. 29–35). https://doi.org/10.2514/1.45525

Campillo Labrandero, M., Martínez González, A., García Minjares, M., Guerrero Mora, L., & Sánchez Mendiola, M. (2019). *Desempeño académico y egreso en 25 generaciones de estudiantes de la Facultad de Medicina de la UNAM*. https://doi.org/10.1016/j.edumed.2019.05.003

Devasia, T., Vinushree, T. P., & Hegde, V. (2016). Prediction of students performance using Educational Data Mining. *Proceedings of 2016 International Conference on Data Mining and Advanced Computing, SAPIENCE 2016*, 91–95.

Gatica-Lara, F., Mendez-Ramirez, I., Sanchez-Mendiola, M., & Martinez-Gonzalez, A. (2010). Variables asociadas al éxito académico en estudiantes de la Licenciatura en Medicina de la UNAM. *Revista de La Facultad de Medicina de La UNAM*, *53*(5), 9–18.

Koedinger, K. R., D'Mello, S., Mclaughlin, E. A., Pardos, Z. A., & Rosé, C. P. (2015). Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(4), 333–353.

Mhetre, V., & Nagar, M. (2018). Classification based data mining algorithms to predict slow, average and fast learners in educational system using WEKA. In IEEE (Ed.), *Proceedings of the International Conference on Computing Methodologies and Communication, ICCMC 2017* (pp. 475–479). https://doi.org/10.1109/ICCMC.2017.8282735

Nieto Domínguez, D. M., Varela Ruíz, M., & Fortoul, T. I. (2003). Aprendizaje: género, escuela y rendimiento académico. *Revista de La Educación Superior*, *32*(125), 53–66.

Urrutia Aguilar, M. E., Ortiz Leon, S., Fouilloux Morales, C., Ponce Rosas, E. R., & Guevara Guzman, R. (2014). El rendimiento académico en el primer año de la carrera de médico cirujano: modelo multivariado explicativo. *Gaceta Médica de México*, *150*(Suppl 3), 324–330.

Yukselturk, E., Ozekes, S., & Kılıç Türel, Y. (2014). Predicting Dropout Student: An Application of Data Mining Methods in an Online Education Program. *European Journal of Open, Distance and E-Learning*, *17*(1), 118–133. https://doi.org/10.2478/eurodl-2014-0008

Using YouTube Analytics to Measure the Effectiveness of Instructor-Generated Video in Online Courses

Matt Farrell
Fanshawe College
mfarrell@fanshawec.ca

ABSTRACT: As institutions shift to remote learning models, instructor-generated video has become an important tool for educators; however, there are significant time-costs associated with video creation. This poses a challenge for instructors who already face increasing demands on their time. How, then, should an instructor approach the creation of video resources in a manner that is optimal for instructors and students? This project explored two potential measures of engagement and retention for online course videos. The YouTube analytics dashboard was used to derive indicators that accounted for class size. These indicators were applied to four different categories of instructor-generated videos in attempt to assess the engagement and retention of each video type. Results identify unique usage patterns across videos. Content delivery videos were viewed more frequently than other types, while introductory videos are viewed more often at the beginning of a term. These data can provide valuable information for educators seeking to assess the effectiveness of their video content, and to optimize their video creation resources.

Keywords: Learning analytics, video analytics, video usage, learner experience design

1 VIDEO LEARNING ANALYTICS

Video learning analytics provides important information about the way students interact with videos in online courses (Mirriahi & Vigentini, 2017). Additionally, Sinha, Jermann et al., (2014) note that the manner in which a student interacts with a video may give insight into the video's level of difficulty or its relevance to assessments. This information can assist instructors seeking to effectively incorporate video into online courses.

1.1 YouTube Analytics

The YouTube video studio provides video analytics features for comparing the reach and impact of videos posted by creators. Many of the indicators are designed with a focus on marketing and revenue generation and are not directly applicable for educators. Draus et al. (2014) used YouTube's analytics dashboard for their analysis of instructor-generated videos, however the raw numbers illustrating minutes-watched do not provide any contextual information about differences in video length or class sizes.

2 STUDY PARAMETERS AND INDICATORS

This study was conducted by an instructor at a liberal arts community college. The courses under study are all elective social sciences courses within the General Education (Gen-Ed) curriculum. The courses employ a standard layout and design, and all courses use four types of instructor generated video resources: 1) Weekly overview videos, 2) Content delivery, 3) Assignment instructions, and 4) Assignment feedback. The indicators for comparison were derived with data from online courses across 6 terms from 2018-2020.

2.1 Indicators

The first indicator, measuring engagement, is *Views Per Student* (VPS). This indicator compares the number of video views to the total class size. A video watched by half of students would suggest a low level of engagement (i.e. 0.5 VPS). Conversely, a score greater than one (i.e. 1.5 VPS) denotes a video watched by all students with several students viewing more than once. This could suggest an especially salient video or difficult topic.

The second indicator is *Average Percentage Viewed* (APV). This indicator, a measure of retention, reports the difference between YouTube's watch time statistics with the total video duration. The measure gives the mean time spent on each video. A low APV score suggests students, on average, are watching only a portion of the video. Similarly, an APV approaching 1 suggests students are watching the full duration of the video.

3 RESULTS

Table 1 displays the results of the study.

Table 1: Results

Video Type	No. of Videos	Total Views	VPS	APV
Weekly Overview	111	1285	.30	71.36
Content Delivery	328	11779	.87	62.59
Assignment Instructions	23	415	.43	70.49
Assignment Feedback	13	240	.47	43.01

As the data suggests, there are differences among the engagement and retention for the different video types. For example, weekly overview videos receive the lowest engagement, yet they seem to have the highest retention. Additionally, content-delivery videos received the most engagement.

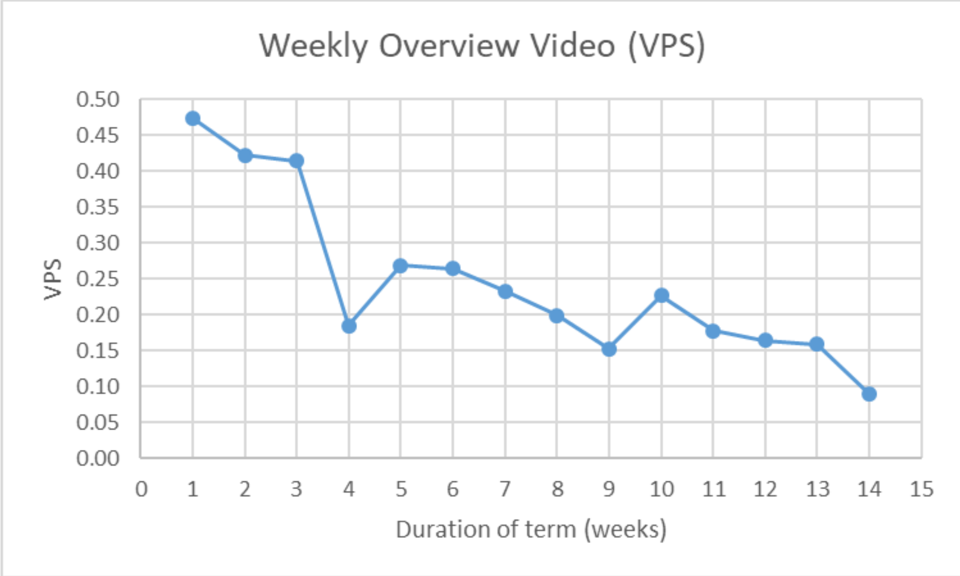
4 DISCUSSION

In addition to the categorical comparison, the two indicators used in this project may also provide actionable insight for instructors at a micro level. Two examples are discussed below.

4.1 Weekly Overview Videos.

Figure 1 illustrates the decline in VPS across the duration of a term. Fewer students watched the weekly overview videos as the term progressed.

Figure 1: Weekly Overview Video Engagement



4.2 Content difficulty

Interestingly, when examining the top four videos ranked by VPS, the same two videos appeared twice, with multiple students viewing the videos more than once. This could suggest a particularly challenging or salient topic. Additionally, the top 5 APV scores were over 100%, indicating that individual students were rewinding and re-watching parts of the videos. Again, this could direct instructors' attention to a challenging topic or incomplete presentation of the course material.

5 NEXT STEPS

This project examined the effectiveness of instructor-generated videos according to the two proposed metrics. Results suggest that instructors might be advised to redirect video creation time to specific video formats, and in some cases, to refine their presentation of certain topics. It should be noted, however, that the raw data don't give a complete picture of the learning experience. Further research could supplement the viewing statistics with survey data to elicit a student perspective of video effectiveness.

REFERENCES

Draus, P., Curran, M., & Trempus, M. (2014) The influence of instructor-generated video content on student satisfaction with and engagement in asynchronous online classes. *Journal of Online Learning and Teaching*, 10(2), 240–254

Mirriahi, N., & Vigentini, L. (2017). Analytics of Learner Video Use. In C. Lang, G. Siemens, A. Wise & D. Gašević (Eds.), *Handbook of Learning Analytics*. http://dx.doi.org/10.18608/hla17.022

Sinha, T., Jermann, P., Li, N., & Dillenbourg, P. (2014). Your click decides your fate: Inferring Information Processing and Attrition Behavior from MOOC Video clickstream Interactions. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 3–14, October 25-29, 2014, Doha, Qatar. https://arxiv.org/abs/1407.7131

Group Perception vs. Group Reality: Exploring the Fit of Self-Report and Log File Data in the Process of Collaboration

Clara Schumacher[1], Natalia Reich-Stiebert[2], Jakub Kuzilek[1], Marc André Burchart[2], Jennifer Raimann[2], Jan-Bennet Voltmer[2], and Stefan Stürmer[2]

[1]Humboldt Universität zu Berlin, [2]FernUniversität in Hagen
clara.schumacher@hu-berlin.de

ABSTRACT: In distance education, online collaboration between students working on joint projects is increasingly promoted and shows beneficial effects on both individual and group outcomes. Besides the positive effects, collaboration could also be accompanied by different levels of student engagement and lead to different degrees of collaborative behavior. However, whether students' actual collaborative behavior and their subjective perception of their collaboration is in line, is an open issue. This poster presents a study conducted at a German distance higher education institution, investigating students' perceptions of task-related communication in comparison to their actual collaborative time on task, as indicated by their trace data in the digital learning environment. While the results indicated that groups spending low and high time on collaboration perceived a similar level of collaboration, this was not represented in their actual collaborative time on task. Future research should investigate the quality of the contributions to collaboration as well as the effects of socio-emotional group processes.

Keywords: collaborative learning, distance higher education, learning analytics

1 INTRODUCTION

Collaboration and communication in diverse (virtual) teams are considered to be crucial skills in the 21st century (Binkley et al., 2012). Thus, students are increasingly being requested to work together in collaborative groups. In fact, current digital learning environments implemented in higher education strive for facilitating group-based processes such as discussions or collaborative project work. To determine whether these forms of collaborative learning are effective, analyses of students' contributions to collaborative learning are relevant. These analyses can be based upon self-report data that reflect learners' perceptions of their learning processes but can also be biased by inaccurate recall or estimations and distortions (Winne, 2017). In that regard, log file data can add a different perspective and contribute to a more holistic picture (Binkley et al., 2012). These additional insights enable teachers to offer immediate interventions to the groups; and by making group processes more visible, the groups' awareness of their collaboration could be fostered (Kirschner et al., 2014). Such support is particularly relevant because collaborative learning requires a great deal of effort of the group members, as not only the individual but also the learning processes of the whole group need to be regulated (Järvelä et al., 2016). These activities take place in different but reciprocal phases and include, for instance, task-related communication or the coordination of group activities (Han & Bayerlein, 2016). While some research has been carried out using self-report data to investigate the collaboration taking place in virtual groups, there remains a paucity of evidence on how these data fit students' actual collaboration. In order to address this open issue, we investigated whether groups in a virtual learning environment differed regarding the collaborative time on task as indicated by log data, and the group members' perception of the collaboration over time.

2 METHODS

2.1 Participants and Design

The data of N = 1917 freshmen (73.4% female) were analyzed in the present study, which was conducted in a lecture on academic working in psychology at a German distance university. After the students were randomly assigned to groups of eight, their central task was to summarize a research paper using the wiki activity in Moodle over a period of six weeks. At the end of each week, students participated in a survey that captured their perception of their group collaboration. To assess actual collaborative behavior, students' log file data were utilized. Based on the log data, the groups without at least seven students active over the period of the task-related group work have been filtered out and 213 groups (of initial 357 groups) remained for further analyses.

2.2 Measures

Perceived task-related communication. To examine students' subjective perception of their collaboration, group members were asked to indicate how much time they had spent on task-related communication (one item scaled to a range from 0 to 1, higher values = more task-related communication).

Actual collaborative time on task. Given T as the time period allocated for solving the group task and T_{gw} as the total time that at least two group members worked together on the task, the measure can be defined as the normalized collaborative time on task as: T_{gw}/T (value range from 0 to 1).

3 RESULTS

To investigate the research question, the sample was divided into two types of groups spending low and high collaborative time on task using the median value (see Figure 1). A repeated-measures MANOVA with the dependent variables *perceived task-related communication* and *collaborative time on task*, the within-subject factor *time*, and the *type of group* as between-subject factor showed a significant main effect for *type of group* (Wilk's Λ = .401, F(2,210) = 156.52, p < .001, η^2 = .599), a significant main effect for *time* (Wilk's Λ = .187, F(10,202) = 88.05, p < .001, η^2 = .813), and a significant interaction effect for *time* × *type of group* (Wilk's Λ = .552, F(10,202) = 16.37, p < .001, η^2 = .448).

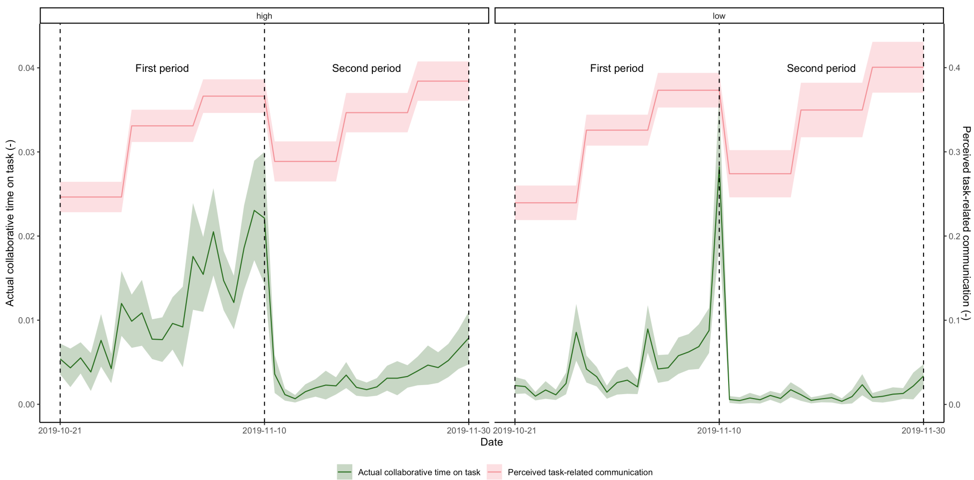


Figure 1: Perceived task-related communication and actual collaborative time on task of groups with high vs. low actual collaboration time.

Univariate ANOVAs indicated significant differences for perceived task-related communication over time (F(4.51,951.97) = 55.34, p < .001, η^2 = .208), and for collaborative time on task over time (F(3.09,652.63) = 124.63, p < .001, η^2 = .371). Yet, the significant interaction effect for time × type of group was only found for actual collaborative time on task (F(3.09,652.63) = 29.21, p < .001, η^2 = .122), but not for perceived task-related communication (F(4.51,951.97) = .545, p = .724, η^2 = .003).

4 DISCUSSION AND CONCLUSION

Based on these results, the relationship between the perceptions of relevant collaborative processes such as task-related communication and groups' actual collaborative time on task seems only limited. As indicated in Figure 1, in the first period of group work at least the shape of the curves of perceived and actual collaboration are comparable, but the amount of time spent on the task-related communication seemingly differs between the two types of groups, and might be overestimated particularly by the groups spending less time on collaboration. In addition, both the perceived and actual time spent on the task increased towards the first deadline of the assignment (see dashed line in Figure 1). Thus, log file data can be considered as a valuable additional source for investigating collaboration but require further empirical evidence. However, the current analyses focused only on task-related measures, without considering socio-emotional processes or learner characteristics and their impact on collaboration (Panadero & Järvelä, 2015). Furthermore, collaboration that took place outside the digital learning environment was not investigated which might have led to different results on collaborative time on task. In addition, the relation of collaborative time on task and group performance needs to be analyzed further. Future research should explore the quality of individual contributions to collaborative tasks and not just the time spent on a task. Finally, the students' perceptions of all relevant processes (e.g., socio-emotional, coordination, regulation) as well as other behavioral and quality indicators should be examined for their potential to predict group performance.

REFERENCES

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 17-66). Springer.

Han, S. J., & Beyerlein, M. (2016). Framing the effects of multinational cultural diversity on virtual team processes. *Small group research*, *47*(4), 351-383.

Järvelä, S., Kirschner, P. A., Hadwin, A., Järvenoja, H., Malmberg, J., Miller, M., & Laru, J. (2016). Socially shared regulation of learning in CSCL: Understanding and prompting individual-and group-level shared regulatory activities. *International Journal of Computer-Supported Collaborative Learning*, *11*(3), 263-280.

Kirschner, P. A., Kirschner, F., & Janssen, J. (2014). The collaboration principle in multimedia learning. In R.E. Mayer (Ed.), *The Cambridge handbook of multimedia learning (2 ed., pp.* 547-575). Cambridge University Press.

Panadero, E., & Järvelä, S. (2015). Socially shared regulation of learning: A review. *European Psychologist*, 20, 190-203. doi:10.1027/1016-9040/a000226

Winne, P. H. (2017). Learning analytics for self-regulated learning. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of learning analytics* (1 ed., pp. 241-249). Society for Learning Analytics Research.

Determining the Interpretability and Viability of Structural Topic Modeling to Analyze Diversity, Equity, and Inclusion Simulation Data

Sydney Dell, Aria Eppinger, Joshua Littenberg-Tobias
Teaching Systems Lab, Massachusetts Institute of Technology
ariaepp@mit.edu

ABSTRACT: Given the increased demand for diversity, equity, and inclusion (DEI) professional development tools, massive open online courses (MOOCs) with digital clinical simulations (DCS) can provide an effective method to teach educators DEI. Identifying patterns in DCS data reveals participant understanding and ways to improve learning through MOOCs; however, the conventional qualitative analysis technique using human raters is time consuming. Our purpose is to determine if structural topic modeling (STM), an unsupervised machine learning technique, is an effective method for analyzing DCS data. Using a random sample of responses, we evaluated whether the most prevalent topic identified by the model for a document matched the human raters' judgments about its contents. The agreement rate between human raters and the model was 65%, which is significantly better than chance. Percent agreement did not vary based on length of the response, but keywords and topic specificity were related to higher agreement rates. Building STMs with more concise topics will make STM a viable approach to analyzing DCS data and improving MOOC learning.

Keywords: Structural Topic Modeling, Machine Learning, Equity, Education

1 BACKGROUND

The growing importance of diversity, equity, and inclusion (DEI) in educational settings brings about a growing need for professional development resources for education staff (Kapila et al., 2016). With asynchronous, self-paced content, massive open online courses (MOOCs) can provide readily available educator training on many topics, including DEI (DeBoer et al., 2014). Digital clinical simulations (DCS), which prompt course participants to engage with material in a mock professional setting, represent a key tool in effectively teaching through MOOCs (Borneman et al., 2020).

However, it is difficult to analyze attitudes present in participant responses and to what extent they were altered by the DCS. Traditional qualitative analyses of open, text-based responses involve a human rater who identifies topics in the responses to gauge user attitudes and understanding of the material (Borneman et al., 2020). By identifying topics, researchers can analyze the prevailing participant mindsets and misconceptions to further improve the MOOC.

Structural topic modeling (STM), an unsupervised machine learning technique for identifying subject matter in text responses, can provide an efficient alternative to analyzing DCS data (Roberts et al., 2014). Specifically, using STM to analyze DCS responses can provide a better understanding of how to effectively teach DEI. However, there is limited research on whether unsupervised models can accurately identify patterns within DCSs, such as participants' attitudes and understanding. We will use an STM to assign topics to participant responses, use human raters to evaluate the interpretability of the topic assigned by the model, and use agreement between the topics identified by the model and the human rater to evaluate the interpretability of the STM-identified topics.

2 METHODS

We analyzed open-ended survey responses to *Jeremy's Journal,* a DCS part of a MOOC on equity in education in Spring 2020 (N=865). Acting as Jeremy's middle school English teacher, participants help Jeremy with academic and at-home struggles. Designed to teach the "Equality-Equity" framework (Milner, H.R., 2016), the DCS asks participants to share their initial beliefs, to role-play as if they were Jeremy's teacher, and to reflect on their teaching mindsets (Borneman et al., 2020).

STM modeling was performed on the responses (N=12,913), and we extracted 15 topics that reflect how participants apply equity and equality mindsets in response to Jeremy's changing situation. To measure the interpretability of the STM results, we randomly sampled roughly 10% of all responses (N=1,207) to be validated by human raters (N=4). Using a method adapted from Chang et al., 2009, raters were given participant text responses and selected one of four possible topics that best categorized the response. Raters could also select "None," suggesting that none of the presented topics represented the text response. We calculated the rate of agreement between the topic selected by the human rater and the topic determined to have the highest probability by the STM. The percent agreement was further broken down by question and by topic. Additionally, we analyzed associations between model confidence and factors, such as percent agreement and response length. We used the model's theta value, which represents its confidence in identifying a topic in a participant response.

3 RESULTS

Human raters agreed with the model's predicted topic 65.2% of the time, which was significantly greater than identifying topics at random (25%) ($t=225.77$, $df=1203$, $p<0.001$). The average percent agreement broken down by rater ranged from 55.6% to 69.2% with a relative standard deviation of 7.7%, suggesting there was little difference between human raters. When the percent agreement was broken down by question, reflection prompts had a lower percent agreement.

Additionally, the average percent agreement varied widely based on the topic identified. More concise topics and those specific to a DCS question had a higher average percent agreement than more broadly defined topics. For example, a topic such as "Doctor's note and school policy" (83.7%) with exclusive keywords like "doctor," "note," and "policy" had higher average agreement than topics such as "Jeremy's focus in class" (42.9%) and "Evaluating Jeremy's Understanding" (45.5%) where keywords like "Jeremy," "quiz," and "class" may have been indicative of multiple topics.

Using a logistic regression, we analyzed the relationship between percent agreement and model confidence (theta value). We found that higher theta values were associated with higher percent agreements (coef=4.1449, s.e.=0.4456, $p<0.0001$). Furthermore, while response length varied widely, there was a significant but very small correlation between the length of the answer and the model's confidence in its topic prediction ($r = -0.0433$, $p < 0.001$), suggesting that the STM does not perform better on longer documents.

4 DISCUSSION

Overall, we determined that the majority of the human raters' judgements matched the topics identified by the STM model (65.2% agreement). Further analyses suggest possible ways to improve model accuracy. One improvement could be using more narrowly defined topics that were more easily identified by the model. Given that accuracy varied by question type, writing DCS prompts that aid STM topic identification would improve accuracy. With continued work, STM represents a viable method to analyze topics in open-text DCS responses and to improve the effectiveness of MOOC in teaching online DEI professional development.

REFERENCES

Borneman, E., Littenberg-Tobias, J., & Reich, J. (2020). Developing Digital Clinical Simulations for Large-Scale Settings on Diversity, Equity, and Inclusion: Design Considerations for Effective Implementation at Scale. *Proceedings of the Seventh ACM Conference on Learning @ Scale*, 373–376. https://doi.org/10.1145/3386527.3405947

Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. *Proceedings of the 22nd International Conference on Neural Information Processing Systems*, 288–296.

DeBoer, J., Ho, A. D., Stump, G. S., & Breslow, L. (2014). Changing "Course": Reconceptualizing Educational Variables for Massive Open Online Courses. *Educational Researcher*, *43*(2), 74–84. https://doi.org/10.3102/0013189X14523038

Kapila, M. Hines, E., and Searby, M., Why diversity, equity, and inclusion matter. (2016). Retrieved from https://independentsector.org/resource/why-diversity-equity-and-inclusion-matter/

Milner, H.R. (2016). *Start where you are, but don't stay there: Understanding diversity, opportunity gaps, and teaching in today's classrooms*. Harvard Education Press.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., Albertson, B., & Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, *58*(4), 1064–1082. https://doi.org/10.1111/ajps.12103

Developing Effective Visualizations to Understand and Scaffold Collaborative Textual Practices

Anisha Singh
University of Maryland
asingh8@umd.edu

Chenxi Liu
University of Maryland
lcx@umd.edu

Revati Naik
University of Maryland
revatin@umd.edu

Philip J. Piety
University of Maryland
ppiety@umd.edu

ABSTRACT: This **poster** will describe and compare different visualizations for collaborative writing using cloud-based platforms (e.g., Google Docs). The aim is to develop effective visualizations for a Learning Analytics Dashboard (LAD) to understand the processes undergirding cloud-based group writing and to support team textual practices. Theoretically situated in the literatures on Collaborative Learning Analytics, collaborative sense making, and social annotations, the visualizations will provide valuable and varied insights to researchers, educators, and students. We will be testing the visualizations and the final LAD with undergraduate students, instructors, and researchers from the fields of learning sciences and educational psychology.

Keywords: Cloud-based writing; collaborative learning analytics; Google docs; social annotation

1 CLOUD-BASED COLLABORATIVE TEXTUAL PRACTICES

Collaboration and communication are critical 21st century skills for the knowledge economy (Trilling & Fadel, 2012). Students must develop these competencies to be successful in teams. In today's information-rich world, these skills are also essential for collaborative sensemaking, wherein diverse people engage with complex information to develop a shared understanding (Ntuen et al., 2006). This understanding is often externalized into a textual product. Such collaborative textual practices are becoming increasingly widespread through adoption of cloud-based tools, such as Google Docs and other tools that allow commenting called *social annotation* (Zhu et al., 2020). However, there is a paucity of research on how students engage in group authoring and ways in which educators can support them in virtual environments in large part due to lack of synthesized data.

1.1 Collaborative Learning Analytics: Visualizations in LAD

Collaboration Learning Analytics (CLA) is an emerging subset of the larger field of learning analytics that is gaining popularity due to its capability to capture authentic experiences through data collection at scale (Piety et al., 2020). The data collected through the CLA can be synthesized and presented to stakeholders (researchers, educators, students) through Learning Analytics Dashboards (Verbert et al., 2013).

1.1.1 Objective

This poster describes and compares different visualizations that would be employed in a LAD to provide stakeholders with actionable insights on collaborative writing in virtual environments. The different types of visualizations under development will capture individual and team-level engagement and collaboration as students use cloud-based platforms in small groups. For this poster, we will be working with Google Docs and collecting data through its API.

2 METHOD

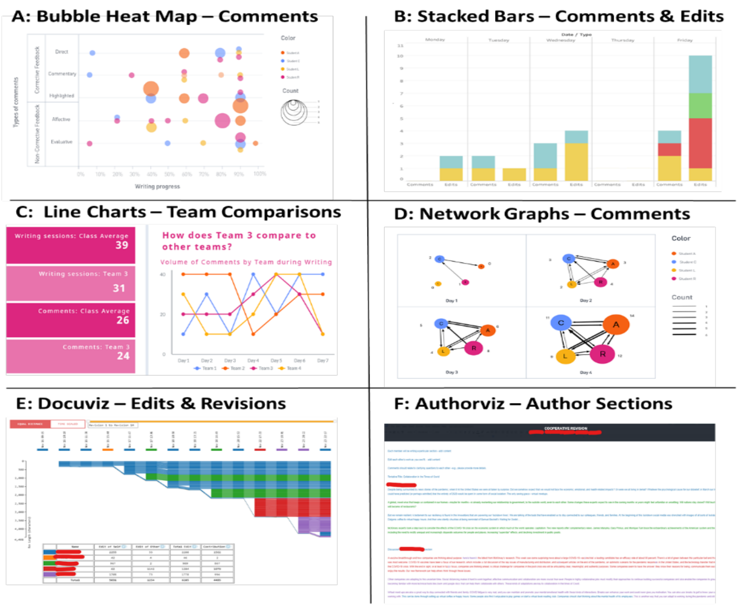
As part of a NSF project (Award #1915563), undergraduates will work in groups to co-author a document in Google Docs. During and after the writing assignment, students will be given feedback using visualizations in a dashboard. Instructors and an inter-disciplinary team of researchers will also have access to this data. The data on the edits, comments, and responses by each team will be extracted using the Google Docs API. This data will be converted into different visualizations targeting students, instructors, and researchers. Several visualization types will be tested. In a mixed method exploratory design, the stakeholders will complete surveys and follow-up interviews about the effectiveness of the visualizations. We would also measure the impact the visualizations in terms of actions taken by students as a direct result of the visualizations.

3 UNPACKING COLLABORATIVE WRITING THROUGH VISUALIZATIONS

Examples of the visualizations under development are presented in Figure 1. The last two visualizations DocuViz and AuthorViz (Wang et al., 2015), do not convey information about comments and communication patterns, which is a critical parameter in cloud-based collaborative writing. Further, they are visually complex whereas the proposed visualizations (A to D) have been designed to be informative to the stakeholders. Visualization A provides information on both the quantity and quality of comments/annotations. Targeted to educators and researchers, it can be used for formative feedback and understanding team engagement. Work progress depicted on the y-axis can be set based on the word limit of the assignment. Visualization B combines information about comments and edits over the week. Meant for students, educators, and researchers, it can help collaborators keep the team accountable and for educators to determine level of progress. For researchers it gives high-level information at a glance. Visualization C has been designed primarily for students and educators. It compares performance across teams and would update daily. Educators can use it to identify teams that need support to accelerate progress. Visualization D are network graphs depicting patterns of exchange between the collaborators. The networks can accommodate larger teams, and the data can be used to determine centrality measures, which would identify team dynamics, such as students who control the flow of information and those who are the most influential. This is targeted to researchers who study teams and communication flows.

4 SIGNIFICANCE AND FURTHER RESEARCH

The visualizations in the LAD will provide data-rich, actionable insights to support group-authoring in cloud-based environments. Given the widespread and ever-increasing use of collaboration tools supporting textual production as learning and professional work move to remote settings, this is both timely and critical. The testing of the LAD and feedback about its visualizations from stakeholders will result in the design of more effective versions in the future. In terms of additional visualization types, epistemic networks of comments and responses can be added to get qualitative insights on communication patterns. Finally, the development of the visualizations for the LAD described here for Google Docs can be expanded to other collaborative cloud-based textual tools.



G. Visualization Comparisons

Information Feature	A	B	C	D	E	F
	Visualization Approach					
Comments quantity by collaborator	X	X		X		
Comment Quantity by teams			X			
Type of comments by collaborators	X					
Quantity of edits by collaborator		X			X	X
Number of writing sessions by teams			X			
Location of comments in the document	X					
Location of edits in the document					X	X
Temporal dimensions of comments/edits		X	X	X	X	
Relation of comments to edits		X				
Communication patterns				X		

Figure 1: Visualizations for understanding collaborative writing

REFERENCES

Piety, P.J., Hansen, D., Bonsignore, E.M., Hickey, D.T. (2020). Collaboration from communication to coordination: Analytics from a project designed to teach collaboration [Manuscript in preparation].

Ntuen, C.A., Munya, P. & Trevino, M. (2006). An approach to collaborative sensemaking process. In Proceedings of the International Command and Control Technology Symposium (ICCRTS), Cambridge, UK.

Trilling, B., & Fadel, C. (2012). 21st century skills: Learning for life in our times. John Wiley & Sons.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. American Behavioral Scientist, 57(10), 1500-1509.

Wang, D., Olson, J. S., Zhang, J., Nguyen, T., & Olson, G. M. (2015, April). DocuViz: Visualizing collaborative writing. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 1865-1874).

Zhu, X., Chen, B., Avadhanam, R. M., Shui, H., & Zhang, R. Z. (2020). Reading and connecting: Using social annotation in online classes. Information and Learning Sciences.

Comparing Natural Language Processing Methods for Text Classification of Small Educational Data

Tanner Phillips, Asmalina Saleh, Krista D. Glazewski, Cindy E. Hmelo-Silver
Indiana University
tanphil@indiana.edu, asmsaleh@indiana.edu, glaze@indiana.edu, chmelosi@indiana.edu,

Seung Lee, Bradford Mott, James C. Lester
North Carolina State University
sylee@ncsu.edu, bwmott@ncsu.edu, lester@ncsu.edu

ABSTRACT: Over the past decade, new natural language processing techniques have been developed that show promise when applied to educational data. Although these methods can be effective, little work has been done to measure the comparative strength of these methods when applied to small data sets. This poster presents an analysis of student chat from a collaborative game-based learning environment. Natural language processing techniques were used to attempt to match human coding of 2877 student chat messages. Findings showed that simple feature engineering methods such as latent semantic analysis outperformed neural networks, which may suggest that it is not appropriate to apply neural networks to the small data sets often found in educational settings.

Keywords: Computer-supported collaborative learning, conversational agents, small data, text classification.

1 INTRODUCTION

Computer supported-collaborative learning (CSCL) environments can offer spaces for students to participate in socially supported education. However, CSCL environments require continued monitoring by instructors and often take place in brick-and-mortar classrooms (Kapur & Kinzer, 2007). One way to address the need for teacher guidance is through conversational agents as a method for augmenting teacher interaction. Conversational agents can monitor student performance in CSCL and offer guidance when students get stuck or go off-task, alerting teachers when human intervention might be needed. However, no consensus has been reached in the learning analytics community as to what method best accomplishes the task of understanding student utterances—the first step in creating a conversational agent. There is high variability in the methods utilized to understand student utterances. They include the utilization of hard-coded or pre-trained linguistic models (Jung & Wise, 2020; Kovanović et al., 2018; Pennebaker et al., 2007), statistical dimensionality reduction techniques (Kovanović et al., 2018; Stone et al., 2019; Vytasek et al., 2019), and neural network models (Fiacco et al., 2019; Stone et al., 2019). In most studies, only one type of model is considered, making comparisons between these models difficult. In this study, we address this issue by comparing a variety of methods for understanding student utterances.

2 METHODS

The student utterances analyzed in this study were gathered from a collaborative game-based learning environment designed to teach students to 14 about environmental science (N=45). In the game, students visit an island in the Philippines where they discover that the tilapia in the local fish farm are sick. Over eight sessions, students visit collected information from characters and objects. Working in groups of four, students use an in-game whiteboard to share their findings with fellow students, engage in group inquiry, and generate hypotheses. They repeat this cycle of collecting data and brainstorming three times before determining why the fish are sick. Throughout the experience, students communicated with each other using an in-game chat feature.

2.1 Data Collection and Utterance Coding

Students' utterances were coded based on the accountable talk and problem-based learning frameworks (Resnick et al., 2018; Saleh et al., 2020). There was a total of eight codes based on student talk (*Agreement*, *Rebuttal*, *Descriptions*, *Hedges*, *Relational*, *Regulation*, and *Questions*). Utterances that could not be coded under these categories were coded as *Other*. The chat data was collected from 45 students between the ages of 12-13. Students worked in 11 groups of four, with one group of five. The students generated a total of 2877 unique chat utterances. We performed standard data cleaning such as removing capitalization and punctuation. However, we did not use certain common NLP practices, such as removing stop words. This was because many utterances contained only a single word (e.g., "no," "hi," or "okay") and removing stop words would have deleted over 20% of the data.

2.2 Classification Methods

The results of ten different NLP models are presented in this paper. These models used one of four methods for feature engineering (LDA, LSA, Pre-trained ELMo word embeddings, untrained word embeddings) and one of three different classification methods (Random Forest (RF), Multinomial Regression (MR), or Long-Short Term (LSTM) Neural Networks). For brevity, several classification methods that performed poorly are not included in this paper including Support Vector Machines and other pre-trained word embeddings such as Word2Vec, GloVe, and BERT. A baseline of 31% was used to measure model accuracy, as this was the size of the largest coding category.

3 RESULTS

Results show that common LDA and LSA machine-learning methods outperformed neural networks at classification (see table 1) when measured against hand coding. MR and LDA performed best overall, the differences were slight between LDA and LSA models. The pre-trained word embedding performed worse than chance, while the free embedding model performed better than chance, but not as well as LDA and LSA models.

Table 1: Results of NLP Classification Models

Feature	Classification	Accuracy	% Above Baseline	Precision	Recall	F-1 Score
LDA	RF	0.425	11.1%	0.424	0.471	0.442
LDA	**MR**	**0.435**	**12.3%**	0.422	0.488	0.440

LSA	**RF**	0.427	11.5%	0.427	0.472	**0.445**
LSA	MR	0.434	12.2%	0.422	0.488	0.440
ELMo	RF	0.234	-7.8%	0.221	0.274	0.246
ELMo	LR	0.269	-4.3%	0.235	0.269	0.278
ELMo	LSTM	0.310	-0.02%	0.309	0.318	0.320
Free Embedding	LSTM	0.380	6.8%	0.355	0.411	0.400

4 DISCUSSION & CONCLUSION

Results support several common assumption of natural language processing, while also suggesting some implications specific to educational data. First, the relatively small data size used in this model does not appear to be large enough for training of neural networks. In the context of K-12 student chat, utterances also often include informal, colloquial, and incomplete sentences, this may help to explain why the ELMo word embedding, which was mainly trained on the Wikipedia corpus, performed so poorly, as it was not familiar with the context it was being asked to analyze. This study also supports the interchangeable usage of LDA and LSA. While all models failed to perform at particularly high accuracy, this study still revealed that what is considered "state-of-the art" in computer-science may not be applicable when dealing with educational data.

REFERENCES

Dascalu, M., Stavarache, L. L., Trausan-Matu, S., Dessus, P., Bianco, M., & McNamara, D. S. (2015). *ReaderBench: An Integrated Tool Supporting both Individual and Collaborative Learning*. 436–437. https://doi.org/10.1145/2723576.2723647

Fiacco, J., Cotos, E., & Rosé, C. (2019). Towards enabling feedback on rhetorical structure with neural sequence models. *ACM International Conference Proceeding Series*, 310–319. https://doi.org/10.1145/3303772.3303808

Jung, Y., & Wise, A. F. (2020). How and how well do students reflect?: Multi-dimensional automated reflection assessment in health professions education. *ACM International Conference Proceeding Series*, 595–604. https://doi.org/10.1145/3375462.3375528

Kapur, M., & Kinzer, C. K. (2007). Examining the effect of problem type in a synchronous computer-supported collaborative learning (CSCL) environment. *Educational Technology Research and Development*, *55*(5), 439–459. https://doi.org/10.1007/s11423-007-9045-6

Kovanović, V., Joksimović, S., Mirriahi, N., Blaine, E., Gašević, D., Siemens, G., & Dawson, S. (2018). Understand students' self-Reflections through learning analytics. *ACM International Conference Proceeding Series*, 389–398. https://doi.org/10.1145/3170358.3170374

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). The Development and Psychometric Properties of LIWC2007 The University of Texas at Austin. *Development*, *1*(2), 1–22. https://doi.org/10.1068/d010163

Saleh, A., Yuxin, C., Hmelo-Silver, C. E., Glazewski, K. D., Mott, B. W., & Lester, J. C. (2020). Coordinating scaffolds for collaborative inquiry in a game-based learning environment. Journal of Research in Science Teaching(57), 1490-1518. https://doi.org/10.1002/tea.21656

Stone, C., Quirk, A., Gardener, M., Hutt, S., Duckworth, A. L., & D'Mello, S. K. (2019). *Language as Thought*. 320–329. https://doi.org/10.1145/3303772.3303801

Vytasek, J. M., Patzak, A., & Winne, P. H. (2019). Topic development to support revision feedback. *ACM International Conference Proceeding Series*, 220–224. https://doi.org/10.1145/3303772.3303816

How K-12 School Districts Communicated During the COVID-19 Pandemic: A Study Using Facebook Data

Joshua M. Rosenberg
University of Tennessee, Knoxville
jmrosenberg@utk.edu

Ha Nguyen
University of California Irvine
thicn@uci.edu

ABSTRACT: The COVID-19 pandemic calls for urgent responses from school districts to allocate resources, develop instructional plans, and retain communication with students and parents. A channel to understand district responses is to examine how school districts communicate via public channels during the pandemic. Yet, understanding the nature of districts' communication when the pandemic was unfolding presents challenges, particularly when the analysis happens at scale. In this paper, we report the results of our use of public data mining, combined with text analysis techniques, to understand how United States school districts communicated about COVID-19. To do so, we compared district Facebook posts from 2019 to 2020 through topic modeling using 50,000 posts, randomly sampled from a corpus of 3,337,147 posts. We also determined how the topics we interpreted differed on the basis of a key school district factor, mean socioeconomic level. This work has implications for understanding K-12 educational institutions' communication about and during COVID-19 and the role of public data and learning analytics research methods for understanding these.

Keywords: public data mining, social media, educational institutions, macro big data

1 BACKGROUND

For communities across the world, the shutdown of schools due to the COVID-19 pandemic disrupted both students' learning and the tenor of daily life for students and their parents and caretakers. In this context, there were differences in how districts communicated with parents, students, and the wider community about their ongoing responses to COVID-19. Exploring this variation through digital—and public—data sources can provide an in-vivo lens (Salganik, 2019) into districts' communication as well as the myriad of ways the pandemic is impacting education.

To explore districts' responses, we took a *public data mining* (Kimmons & Veletsianos, 2018) approach, one that uses public data for research. Given the locally-controlled nature of education in the United States (U.S.), social media may be a channel through which districts communicate about many topics, from those related to enabling students to participate in schooling (e.g., posts about materials and devices) to celebrating students and others in the community. In this way, we employ an approach that is related to the use of *macro* (social media) data sources but it is different in that we focus not on one or a handful of institutions, but all that was accessible to us, and in how we use data from posts over many time points—akin to data collected in learning analytics systems (Fischer et al., 2020). Our research questions, then, are: How did U.S. school districts' communication via social media change in response to the COVID-19 pandemic? And, to understand variation in their communication, how did this differ based upon an indicator of their mean socioeconomic level?

2 METHOD

Our data collection process, including the use of the CrowdTangle platform (CrowdTangle Team, 2020), which provides researchers access to Facebook data, is summarized in Figure 1.

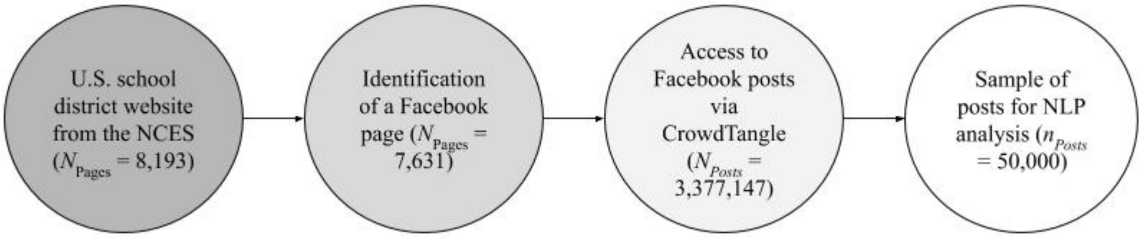


Figure 1: Data collection process used to identify posts from U.S.-based K-12 school districts

We then sampled 50,000 posts, due to the computational challenge of conducting NLP analyses on the entire corpus. The topic models, which were fit using the *stm* R package (Roberts et al., 2019), allowed topic prevalence to vary by year (2019 or 2020) and a measure of socioeconomic status (i.e., the percentage of students participating in a free or reduced-price lunch program, FRLP). To determine the number of topics K, we ran a series of models with K ranging from 10 to 40 and examined several model fit diagnostics, which suggested *K* = 20 indicate high predictive power and high semantic coherence (i.e., high-frequency words from a topic are likely to co-occur). The topic names (left panel, Figure 2) were determined by examining the top 15 high-frequency words and representative posts from each topic, followed by discussions between the authors.

3 FINDINGS

When ranked by frequency, topics about school spirit, help-giving, sports, registration, scheduling, and materials/device were those with the highest average prevalence in the corpus (see Figure 2).

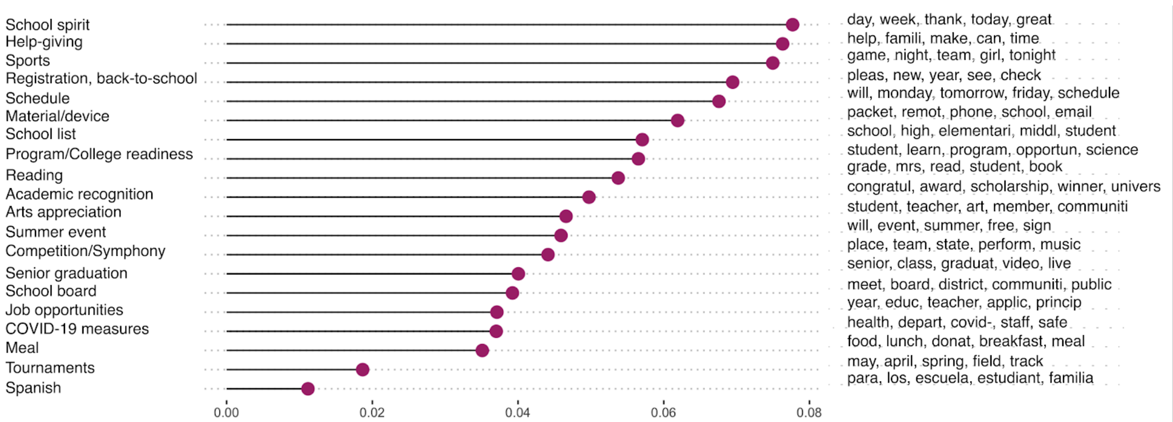


Figure 2: Topic Prevalence, Ranked by Highest-Lowest Proportion, and 5 High-Frequency Words

The topic models also allowed us to examine whether topic prevalence varied by covariates. Results suggest that in 2020, topics about registration/back-to-school, COVID-19, material/device, help giving, and senior graduation were more prevalent (Figure 3). For example, on average, the proportion of topics about material/device pick-up was .06 higher in 2020 than that in 2019.

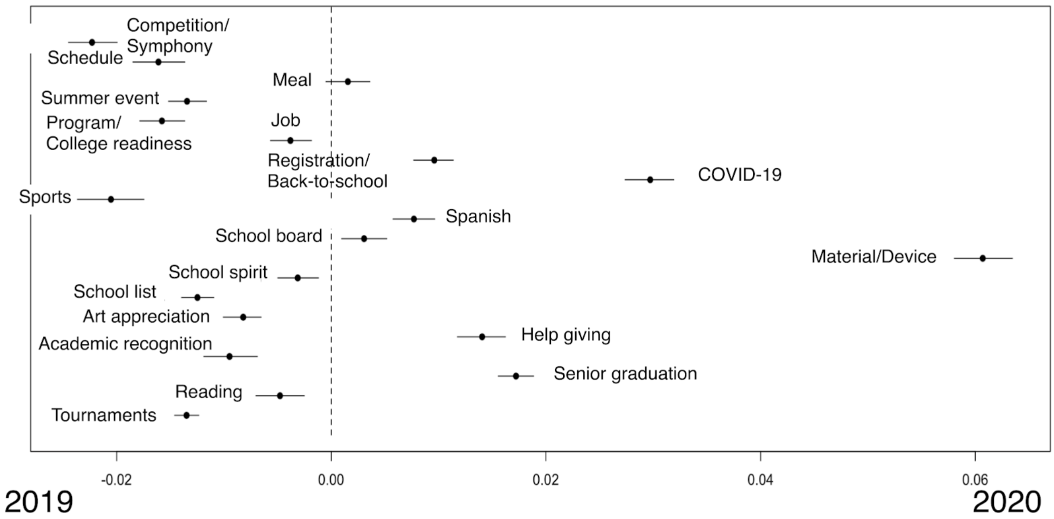


Figure 3: Difference in Topic Prevalence between 2019 and 2020.

Findings from the analysis of differences based upon FRLP suggest that overall, posts from school districts with a lower average socioeconomic level featured higher topic prevalence regarding meals, job posting, and materials/device pickup compared to those with greater socioeconomic resources.

4 DISCUSSION

This work demonstrates how a new data source—social media posts by institutions—can be informative about topics related to teaching and learning at scale. A key methodological feature that enabled this work was the identification of social media links on districts' websites. Indeed, we found that a surprisingly high proportion of districts may be using social media, but this data source has been examined very little in learning analytics research and educational research. Moreover, access to Facebook invites new learning analytics-driven questions (and ethical considerations, especially when posts by individual teachers or students are the focus of study, as they were in the present study), including those about variation in curricula and supports for students.

REFERENCES

CrowdTangle Team (2020). *CrowdTangle.* Facebook, Menlo Park, California, United States.

Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., ... & Warschauer, M. (2020). Mining big data in education: Affordances and challenges. *Review of Research in Education*, *44*(1), 130-160.

Kimmons, R., & Veletsianos, G. (2018). Public internet data mining methods in instructional design, educational technology, and online learning research. *TechTrends, 62*(5), 492-500. doi:10.1007/s11528-018-0307-4

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). Stm: An R package for structural topic models. *Journal of Statistical Software*, *91*(1), 1-40.

Salganik, M. J. (2019). *Bit by bit: Social research in the digital age*. Princeton University Press.

Detecting Learner's Hesitation in Solving Word-Reordering Problems with Use of Machine Learning for Better Precision

Yoshinori MIYAZAKI
College of Informatics, Shizuoka University, Japan
yoshi@inf.shizuoka.ac.jp

Ryosuke BANNO
Faculty of Informatics, Shizuoka University, Japan
banno.ryosuke.16@shizuoka.ac.jp

ABSTRACT: The aim of our present study is to develop a Web application to apprehend when learners hesitate in the process of solving English word-reordering problems. Measurements of learners' hesitation can be an important clue to ascertain learners' understanding level. In the past research, we assumed that the task of classifying study logs into "hesitating" and "not hesitating" labels, and adopted a supervised learning technique in machine learning. Parameters of mouse behavior were used as features for classification. Seeking for its better quality, our newly constructed classifier further incorporated four parameters chosen after scrutiny. The result of the attempt is shown, with the indication of the feasibility to locate where hesitation occurs in the problem-solving process.

Keywords: Word-reordering problem. Mouse trajectory. Occurrence of hesitation. Supervised machine learning. Random forest

1 INTRODUCTION

The advancement of e-Learning systems in educational settings has made it easier to observe study logs. This has helped our research group develop a Web application with which to detect when learners hesitate in the process of solving English word-reordering problems (WRPs). In WRPs learners are asked to make an English sentence from randomly given words, one with a meaning equivalent to the sentence provided in Japanese. This type of problem has been a popular means in Japan to measure learners' English competency, or knowledge of grammatical items such as sentence structure, idioms/idiomatic phrases, and usages. When learners solve WRPs online, they rearrange given words using their mouse's drag and drop (D&D). Using the mouse not only makes it easy for learners to produce answers of WRPs (often more easily than on paper), but it also provides us with an opportunity to record mouse movements and to create a study log. Focusing on the latter feature, we have been developing a Web application that helps us detect learners' hesitation by analyzing mouse behavior. Figure 1 below is two cases of the mouse trajectories—reproduced and visualized in lines—from two correct answers of the same problem. We have hypothesized that the amount of hesitation the learners have experienced and complexity of mouse trajectories created in solving the same problem are deeply connected. If the hesitation can be identified, we will be able to evaluate better the learner's understanding level and their degree of knowledge of the important item(s) tested in WRPs. If a learner's hesitations are found in many of the same type of problems, he/she will be advised to review the problems from that category. Hesitations detected in one learner, or a group of learners, are also useful when teachers work on teaching plans or methods.

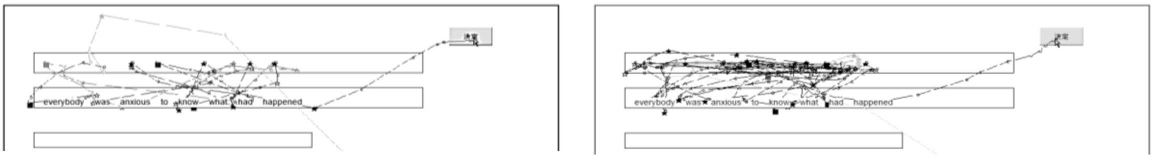


Figure 1: Two correct answers of the same problem with different mouse trajectories (left: solved without hesitation, right: solved with hesitation)

2 WEB APPLICATION

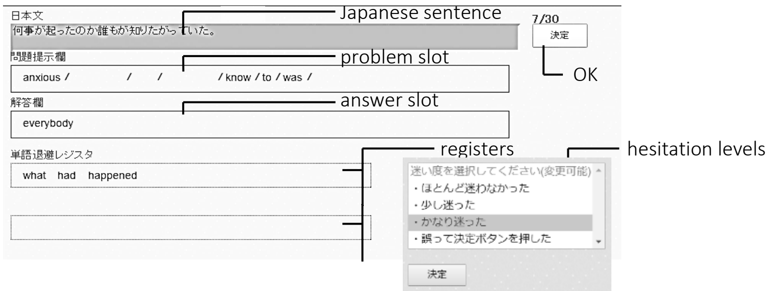


Figure 2: Study module

Our software has Study Module, which requires learners to perform word-reordering tasks by “dragging and dropping” each word into the appropriate position in a sentence, simultaneously recording all the mouse trajectories as well as the timing of D&Ds in answering the problems; and Reproduce Module, intended for both of learners and teachers, to reproduce all the actions in the learners’ mouse trajectories, and analyzes the data from the diverse patterns of the study logs both from the learners’ and problems’ perspectives. In WRPs, learners are asked to reorder given words and the words to be rearranged are given in the “problem slot,” and all the words should be moved into the “answer slot” by D&Ds with the mouse (Figure 2). It is required for learners to press the “OK” button to finish answering, and then they rate their hesitation level in the answers on a four-point scale—”not hesitating,” “a little hesitating,” “hesitating,” or “pushed the button by mistake.” In order to facilitate learners’ performing the tasks, this module has the following functions: 1) Word groupings: an arbitrary number of words can be grouped together by mouse-dragging (rectangular selection) if it is convenient for learners to treat them together, and; 2) Relocation to registers: areas called “registers” are provided as a temporary “shelter” for words, where a set of words can be integrated if it is preferred for learners to organize their ideas. All mouse movements for solving problems are recorded in this module, such as D&D(s), U-turn(s) (the right-and-left or up-and-down mouse movements), and the time used in one treatment of one word, as well as the time elapsed between a particular drop and the click of the next word (D-C time), and standstill time of the mouse.

3 FEATURE EXTRACTION AND EXPERIMENT

In this application, X and Y coordinates of the mouse location at each time step are stored along with its status, with the value of the MouseDown property of the mouse at each timing. From these, the following parameters are computed and used for machine learning as features of mouse trajectories: Answering time / Total distance (of mouse movement) / Average velocity / Longest standstill time / Number of D&Ds / Maximum D&D time / Maximum D-C time / Number of horizontal U-turns / Number of vertical U-turns.

In contrast with the aforementioned nine parameters used in Banno, et al. (2019), we further scrutinized not only the correlation coefficients with the degree of hesitation, but relationship with the reportedly difficult word(s) in each problem. This is how the following four parameters among a dozen of candidates are chosen to seek for better classification: Maximum velocity / Elapsed time until the first D&D is executed / Time elapsed from the first D&D till the end / Total D-C time.

First, the classification problem in machine learning is outlined. A classifier is gained with the learning data consisting of labels and features. After the classifier is created, a label is output by inputting parameter values of the features. The label is set to the degree of hesitation obtained by the answers from the learners. At present, the application asks learners to give the degree of hesitation in solving each WRP. However, the classifier with a high precision will enable the degree of hesitation to be judged only by the study logs created by learners. Implementation has been achieved using scikit-learn (Python) and the random forest algorithm.

To collect label and feature data, 22 students with a variety of majors at a certain university in Japan were asked to solve 30 problems using the WRP system we have developed. From thus collected 660 data, 230 rated ”hesitating” and the same number (of randomly extracted out of 256) rated ”not hesitating” were used (therefore, the baseline is 50%). For the evaluation, we adopted a leave-one-out cross-validation test with ten divisions and produced the ratio of correct answers, precision, recall, and F-measure for performance indices. Table 1 presents the averaged precision, recall, and F-measure for “hesitating” and “not hesitating.” The figures in parentheses are the percentages obtained by the nine parameters (Banno, et al. (2019)). For all criteria, the results of the proposed combination of parameters showed higher correctness by around 1-2%. Let us note that such differences are not considered minor when the figures are getting close to 100%.

Table 1: Precision, recall, and F-measure.

Label	Precision	Recall	F-measure
Hesitating	85.4% (84.2%)	81.7% (80.0%)	83.2% (81.6%)
Not hesitating	82.8% (81.3%)	85.4% (84.1%)	83.7% (82.4%)

4 SUMMARY

In this study, a method was proposed to judge whether learners hesitated in the process of solving WRPs by applying supervised machine learning and using the random forest algorithm. Compared to the previously selected parameters, four of the new ones were incorporated to construct the classifier. As the result of the attempt, gaining higher F-measures by around 1 to 2% was observed.

REFERENCES

Zushi, M., Miyazaki, Y., Norizuki, K. (2015). Analysis of Learners' Study Logs: Mouse Trajectories to Identify the Occurrence of Hesitation in Solving Word-Reordering Problems, The 5th International Conference on Learning Analytics and Knowledge (Poster).

Banno, R., Miyazaki, Y., Zushi, M., Norizuki, K., Hasegawa, Y. (2019). Analysis of Study Logs for Detecting Learner's Hesitation in Solving Word-Reordering Problems, The 94th LET (The Japan Association for Language Education & Technology) Chubu branch conference.

Designing the TEACHActive Feedback Dashboard: A Human Centered Approach

Dana AlZoubi, Jameel Kelley, Evrim Baran, Stephen B. Gilbert, Shan Jiang, Aliye Karabulut-Ilgu
Iowa State University
{dalzoubi, jkelley, ebaran, gilbert, sjiang1, aliye} @iastate.edu

ABSTRACT: Effective facilitation of active learning is key to enhancing student engagement in engineering classrooms. Instructors need opportunities for frequent observation, feedback, and reflection on the use of their active learning strategies, yet there are no validated automated approaches available. We address this need by designing a feedback dashboard, TEACHActive, that leverages classroom analytics from an automated sensing observation system. The TEACHActive dashboard provides feedback on the in-class implementation of various active learning strategies in engineering classrooms. In this poster, we present the initial phases of a human-centered dashboard design process. The human-centered design (HCD) approach includes techniques such as, creating personas, conducting user interviews, and implementing user walk-through sessions. To confirm the practicability of TEACHActive dashboard for further revisions before the actual larger scale (*n*=30) implementation, a small sample of engineering instructors (*n*=5) participated in the prototype design process to identify meaningful attributes associated with the TEACHActive dashboard and shared perspectives and expectations towards its use in their classrooms.
Keywords: feedback dashboard, active learning, classroom analytics, human-centered design.

1 Background

Effective facilitation of active learning in engineering classrooms is key to promoting student engagement (Shekar et al., 2015). The use of automated systems for classroom observation and feedback is growing, yet few studies have integrated a specific classroom pedagogy (Lockyer et al., 2013), and none have addressed it in the context of active learning use in engineering classrooms. There is a critical need for research that links pedagogical theories with in-class practices to determine ways to improve instructors' implementation and facilitation of effective teaching practices (Bodily et al., 2018). We designed the TEACHActive feedback dashboard by leveraging classroom analytics from automated observation to provide feedback on the in-class implementation of various active learning strategies in engineering classrooms. TEACHActive communicates with an existing camera-based automated classroom sensing system, EduSense (Ahuja et al., 2019), which tracks faculty and student proximities and behaviors in a classroom. TEACHActive is designed to transform raw classroom data into meaningful metrics and then further into practical feedback for instructors. TEACHActive dashboard visualizations provide automated feedback for instructors about their facilitation strategies in correspondence with the captured features of interest, including sit vs. stand, hand raises, body position, instructor movement, student vs. instructor speech, and speech acts.

2 TEACHActive Feedback Dashboard Design

Our design approach follows the human-centered design (HCD) principles (Arbas, Maloney-Krichmar, & Preece, 2004). taking into account various human factors of why and how the system and the interface are used. We initiated our HCD approach by first identifying the context for implementation and the instructors as potential users. We then employed various HCD techniques to generate an understanding about instructors' needs, goals, barriers, frustrations, expected outcomes, and experiences.

2.1 Needs Analysis and Persona Development

Our first phase in the TEACHActive dashboard prototype design was creating data-driven user personas. We first built an understanding of potential users through looking at patterns from the findings of a needs analysis that was conducted with engineering faculty. The needs analysis included data collected through a survey (*n*=53) and follow-up semi-structured interviews (*n*=4). Survey questions aimed to gather instructors' perspectives, knowledge, use of active learning strategies in engineering classrooms. The follow-up interviews helped collect data about instructors' teaching experiences, courses taught, specific examples and reasons for active learning implementation in classrooms, classroom management strategies, challenges, support, and desired outcomes. Our thematic analysis of survey and interview data revealed four personas: (a) The Agile, (b) The Seeker, (c) The Planner, and (d) The Feeler (Table 1). All personas share at least one common goal, which is implementing effective active learning strategies to better engage students. Each user persona developed will be shared in the poster session.

Table 1: User Personas

User Persona	Goals	Characteristics (Important factors)	Needs/Support Factors	Frustrations/ Barriers	Expected Outcomes
The Agile	Moving around in class	-Classroom climate -Class size, room structure, seating	-More space -Better technology	-Staying in one spot -Lecturing too much	-More engagement and interaction
The Seeker	-Seeks recognition for good teaching -tracking improvement	-Mobilizing support from faculty & administration -Flexible in changes	-Faculty/peers/administration -Feedback on teaching	-Engaging students -Feedback on teaching	-Progress report -Constructive feedback on ways to improve
The Planner	-Planning good fit activities -Making lectures more interactive	-Balance between lectures and activities -Evidence AL is not a waste of time	-Structured times to integrate activities -Building routine	-Time constraints -Choosing between lectures & activities -Changing plans & class routine	- Proof of progress from one session to another
The Feeler	-Excited about change -More engaging lectures	-Motivated internally by interaction and feedback -Emotionally charged -Fearful & excited	-Reactions from students -Positive reinforcement -Motivation & creativity	-Fearful & nervous about change -Not receiving good feedback	-Proof of progress from one session to another

2.2 Initial Dashboard Prototype Development

Our second design phase was to develop the TEACHActive dashboard prototype iteratively based on the personas created and the features captured by the classroom sensing system. The initial dashboard prototype was designed with Adobe XD and included two main displays: (a) session and (b) progress. The session display included the following metrics: total number of hand raises and their frequency as a function of time, duration of instructor speech, duration of student speech, frequency of instructor vs.

student speech as a function of time, instructor movement patterns, sit vs stand. The progress display included comparison statistics between the session display metrics through bar graphs.

2.3 User Interviews and Walk-Throughs

The user interviews and walk-throughs were carried out on a small scale (*n*=5) to confirm the practicability of TEACHActive dashboard for further revisions before the larger scale (*n*=30) implementation. We conducted thirty-minute semi-structured user interviews with five engineering instructors to understand their perspectives and expectations of the initial dashboard prototype features. In the user walk-throughs, we discussed each of the dashboard metrics and visualizations in terms of their perceived usefulness to identify meaningful attributes associated with the TEACHActive dashboard. Based on the instructor recommendations, we modified the initial TEACHActive dashboard prototype. We will share different versions of the dashboard prototypes in the poster session.

3 Conclusion

The TEACHActive dashboard aims to support instructors' implementation and facilitation of active learning strategies in engineering classrooms using the analytics of classroom sensing data. In this poster, we present our HCD approach for developing the initial dashboard prototype. Next, we will develop further prototypes using the React framework, pilot those with actual classroom video recordings, and create revisions with further instructor walk-throughs.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Grant No. 2021118. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of National Science Foundation.

REFERENCES

Ahuja, K., Kim, D., Xhakaj, F., Varga, V., Xie, A., Zhang, S., Townsend, J., Harrison, C., Ogan, A. & Agarwal, Y. (2019). EduSense: Practical classroom sensing at Scale. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3*(3), 1-26. doi.10.1145/3351229

Abras, C., Maloney-Krichmar, D., & Preece, J. (2004). User-centered design. *Bainbridge, W. Encyclopedia of Human-Computer Interaction. Thousand Oaks: Sage Publications*, *37*(4), 445-456.

Bodily, R., Kay, J., Aleven, V., Jivet, I., Davis, D., Xhakaj, F., & Verbert, K. (2018, March). Open learner models and learning analytics dashboards: a systematic review. *In Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 41-50).

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist, 57*(10), 1439-1459. doi.10.1177/0002764213479367

Shekhar, P., Demonbrun, M., Borrego, M., Finelli, C., Prince, M., Henderson, C., & Waters, C. (2015). Development of an observation protocol to study undergraduate engineering student resistance to active learning. *International Journal of Engineering Education, 31*(2),597-609.

Emotion detector: Employing Machine Learning Models to Identify Students' Emotions in CSCL Environments

Ahmad Khanlari, University of Toronto, a.khanlari@mail.utoronto.ca
Gaoxia Zhu, Cornell University, gaoxia.zhu@cornell.edu

ABSTRACT: While students' emotions may affect their participation in online activities and learning, it is challenging for teachers to grasp an overview of students' emotions in online collaborative learning environments. This study proposes a tool, developed employing Natural Language Processing (NLP) techniques and machine learning models, that detects students' emotions with 78% accuracy. This tool enables teachers to automatically detect students' emotions expressed in their notes written in online collaborative learning environments. The implications and limitations of this study are discussed.

Keywords: CSCL, emotion detection, discourse analysis, machine learning, NLP.

1 INTRODUCTION AND LITERATURE

Emotion is defined as individuals' experiences and responses to the events and social contexts around them (Mulligan & Scherer, 2012). Studies show that both positive and negative emotions affect students' learning and experience (e.g., D'Mello et al., 2014; Pekrun et al., 2017). In Computer-Supported Collaborative Learning (CSCL), interactions between students and their peers, teachers, ideas, technologies and so forth tend to elicit various emotions.

Students' emotions can be analyzed using self-reports, external observation or more technical approaches such as sentiment analysis, facial expression recognition, and biosensors (e.g., Arguedas et al. 2016; Zhu et al., 2020). For instance, Zhu et al. (2020) used IBM Watson's Tone Analyzer to detect the emotional tone of more than 19,000 online written notes into anger, fear, joy, and sadness. Zhu et al. (2019) identified joy, curiosity, neutrality, and challenge in students' collaborative online discussions through manual coding. Curiosity and challenge indicate students' cognitive states during the knowledge construction process. However, manually coding these emotions is tedious work. The present study aims to develop machine learning models to detect these emotions automatically.

2 METHOD

We collected 299 notes from 22 grade 2 students, posted on "Knowledge Forum." Knowledge Forum is an online environment that supports students' collaborative knowledge creation through progressive discourse. The study lasted eight weeks, and the students worked on a mathematics topic: shapes. The students mainly discussed the definition of a shape, shape design, and 2D and 3D shapes in Knowledge Forum. Two researchers discussed all the notes and identified students' emotions expressed in their notes: Joy, Curiosity, Neutrality, and Challenge occurred more frequently in notes and were adopted in the coding. Some other emotions, such as frustration, gratefulness, boredom, and disgust were rarely observed in students' online discussions and were coded as "others."

As shown in Figure 1, the labelled data was first processed to create a clean text by performing word stemming and removing stop words using Python's Natural Language Processing package (NLTK).

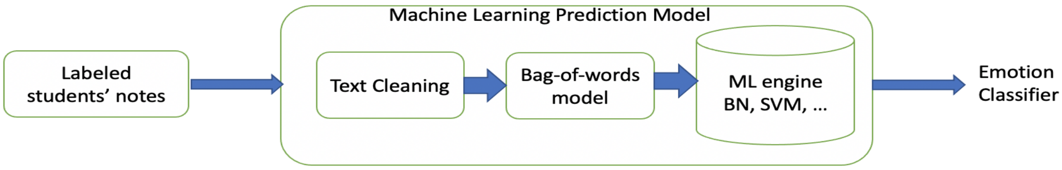


Figure 1: System architecture

In the next step, a Bag-of-words model was created. The Bag-of-words represents a text with a vector that indicates the number of occurrences of each chosen word in the training corpus (Sebastiani, 2002). The data was then randomly split into train and test tests. The train set contained 80% of the labelled data, and the test set included 20% of the labelled data. Different machine learning models were then trained to learn what terms are associated with each emotion. In the final step, the trained models were used to classify the test set.

We compared six state-of-the-art learning algorithms in training the emotion classifier, including Logistic Regression, Bayesian Network, Support Vector Machines, Decision Tree, Random Forest, and K-Nearest neighbours (k-NN). The evaluation metrics of our experiment include accuracy and the confusion matrix. Accuracy is the ratio of correct predictions to the total predictions, while a confusion matrix is the count values and breaks down the correct and incorrect predictions of each class.

3 PRELIMINARY RESULTS

Here, we only report the performance of k-NN, which outperformed other algorithms. The accuracy of the k-NN prediction model is 78%. Figure 2 shows the normalized confusion matrix.

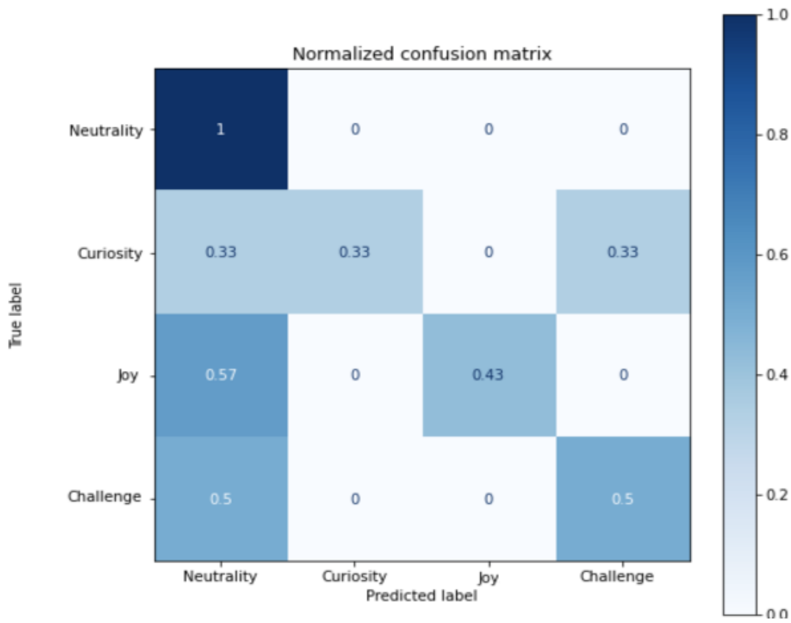


Figure 2: Normalized confusion matrix

As evident in Figure 2, 100% of the Neutrality class is correctly identified by the prediction model. The model was also able to correctly identify 33% of the Curiosity class. However, 66% of the Curiosity notes were identified as Neutral and Challenged. Moreover, as the confusion matrix shows, 43% of Joyful/Satisfied class were correctly identified, while 57% of the Joyful class were identified as Neutral notes. Finally, the prediction model was able to correctly classify 50% of the Challenged class, while the remaining Challenged notes were classified as Neutral. It should be noted that the test set did not include any notes from the "Other" class; therefore, this class is not shown in the confusion matrix.

4 DISCUSSIONS, LIMITATIONS AND FUTURE WORK

Recognizing students' emotions from texts is time-consuming, which requires intensive attention from teachers. The preliminary result of this study is promising since the developed tool could detect students' emotions with 78% accuracy. This tool enables teachers and researchers to automatically detect students' emotions and provide support, when necessary. By focusing on students' curiosity and challenge occurred in CSCL environments, this tool is distinguished from other sentiment analysis tools (see D'Mello, 2017 for a review). One of the applications of this tool is in Massive Open Online Course (MOOC), in which several thousands of students are enrolled. Employing this tool in MOOCs and identifying students' emotions can help instructors take necessary actions in a timely fashion.

However, some limitations need to be addressed in the future. First, the labelled data included only 299 written notes and were imbalanced. There were 173 notes labelled as Neutrality, 70 notes as Joy, 22 notes as Curiosity, 22 notes labelled as Challenge, and the remaining 12 notes labelled as Others. Indeed, the reason that the model correctly identified all the Neutrality notes and many of the Joy notes is that there was a fair number of these notes in the training set. However, as there were only fewer Curiosity notes in the training set, the model was not well-trained to identify this class. We aim to replicate this study using a larger and balanced dataset to develop better classification models.

REFERENCES

Arguedas, M., Daradoumis, T., & Xhafa, F. (2016). Analyzing How Emotion Awareness Influences Students' Motivation, Engagement, Self-Regulation and Learning Outcome. *Educational Technology & Society, 19*(2), 87–103.

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, *29*, 153-170. DOI: 10.1016/j.learninstruc.2012.05.003

D'Mello, S. K. (2017). Emotional learning analytics. In C. Lang, G. Siemens, W. Alyssa, & D. Gašević (Eds.), *Handbook of learning analytics & educational data mining* (pp.115-127). Edmonton, AB: Society for Learning Analytics Research.

Mulligan, K., & Scherer, K. R. (2012). Toward a working definition of emotion. *Emotion Review*, *4*(4), 345-357.

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement emotions and academic performance: Longitudinal models of reciprocal effects. *Child development, 88*(5), 1653-1670.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM computing surveys (CSUR), 34*(1), 1–47.

Zhu, G., Xing, W., Costa, S., Scardamalia, M., & Pei, B. (2019). Exploring emotional and cognitive dynamics of knowledge building in grades 1 and 2. *User Modeling and User-Adapted Interaction, 29*(4), 789-820.

Zhu, G. (co-chair), Teo, C.L. (co-chair), Scardamalia, M., Martin, K., Mohamed, F.,…& Woodruff, E. (2020, June). Emotional and Cognitive Affordances of Collaborative Learning Environments. *The Proceedings of the 14th International Conference of the Learning Sciences* (pp. 382-389). Nashville: International Society of the Learning Sciences (online conference).

Virtual Gemba Analytics for Experiential Learning

Fabiana Santos, Alice Mello, Nikki James
Northeastern University
santos.fa@northeastern.edu, a.mello@northeastern.edu, ni.james@northeastern.edu

ABSTRACT: When developing a project, input from stakeholders is a key to success. In this paper, a Virtual Gemba Walk dashboard for virtual capstone projects is proposed. A Virtual Gemba Walk aligns the expectations of the three stakeholders: student, teacher and industry sponsor through a real-time analytics dashboard that visualizes project indicators, tracks progress and identifies misaligned expectations. This poster presents a proposed interactive dashboard, that leverages data from technology to support the Virtual Gemba Walk process. The dashboard contains key indicators of the capstone project, triggers new Gemba Walks and visualizes feedback from each stakeholders' perspective. The aim is to help students, teachers and industry sponsors to get meaningful feedback for a better chance of project success.

Keywords: Gemba Analytics, Learning Analytics, Stakeholders, Experiential Learning

1 INTRODUCTION

A capstone project is a model of experiential learning that brings real-world projects and sponsors into an academic classroom. A capstone project is a complex pedagogical practice to facilitate, online and remote learning paradigms make it even more complex. This poster showcases how learning analytics can be used to align stakeholders. This alignment is achieved by visualizing issues, combining learning activities with indicators of project management, development and success (Reynolds, 2009). Learning Management Systems (LMS) are traditionally a two-way channel between teachers and students but in an industry engaged, experiential learning program like a virtual capstone project, including the industry sponsor in the virtual learning environment is key. Practera, an experiential learning technology designed to support teacher, student and industry sponsor collaboration (James et al., 2020) captures unique data that can be used to implement the Virtual Gemba Walk process.

Gemba has roots in the Japanese culture and means 'the real place', which in Lean methodology was applied as the place of work where value is created (Petruska, 2018). The Gemba walk is a method used to engage and showcase the current state of a project to leaders, stakeholders, and clients. The proposed Virtual Gemba Walk is a learning analytics driven presentation made by a student team highlighting key project indicators including, project progress and key deliverables.

This poster presents a proposed learning analytics driven Virtual Gemba Walk dashboard prototyped using teacher, student and industry sponsor collaboration data from an experiential learning technology. The Virtual Gemba Walk dashboard supports a remote capstone project team to visualize their progress for teachers and industry sponsors. Additionally, through the Virtual Gemba Walk dashboard Sponsors can identify potential project improvements, flag issues or request further feedback.

2 VIRTUAL GEMBA WALK EXPLAINED

2.1 Data Used

To create the dashboard, the data was acquired from the experiential learning technology. The platform captures data points related to the student use, behavior and completion of project tasks. The student engagement metrics captured include number of project tasks completed, learning content completion, project deliverable submissions, team communications, achievement of badges and points to recognize project performance. Moreover, the technology also captures teacher's interventions, industry sponsor feedback and perspectives on the students' project engagement and project progress. Those indicators were analyzed and used as Key Performance Indicators (KPI) in the Virtual Gemba Walk dashboard.

In addition to the existing data captured by the experiential learning platform, a new data set is proposed to add valuable insight to the Virtual Gemba Walk process. Finally, the Virtual Gemba Walk itself generates feedback that contains a flag indicator based on the status of the project from the perspective of the industry partners, students and teachers and written feedback.

2.2 Purpose

As mentioned previously, a traditional Gemba Walk is a process in which a stakeholder would meet with a team to view the status of the project. In a Virtual Gemba Walk, the industry partner can trigger it manually at any time to assess, evaluate and provide feedback on the project progress. The key purpose of a Virtual Gemba Walk is not only to get effective feedback from the industry Sponsors, but to build trust and develop more effective teams (Gasevic, Dawson & Siemens, 2016).

2.3 Goals & Objectives

The main goal of a capstone project is for students to successfully apply concepts they have learned in the classroom to a real-world project. This is the key factor that can be analyzed, predicted and visualized using a learning analytics driven Virtual Gemba Walk dashboard. However, there are additional factors like work habits, teamwork skills and project quality that are difficult to map using existing data from the technology (Scholes, 2016). The additional feedback provided by students, industry sponsors and teachers during the Virtual Gemba Walk process can provide this additional data not captured by the platform itself.

The overall objective of the Virtual Gemba Walk is student success. In a capstone project this includes processes and indicators that are not just related to learning (Verner, Evanco & Cerpa, 2007). The analysis that produces the Virtual Gemba Walk dashboard uses student success as the focal point. Specifically, it used in a regression model to understand how other variables might affect the level of student success. The regression analysis is done by taking into consideration the correlation between variables and the normalized impact that each variable has on the target variable, student success. The multiple regression is built using the measurements from the student usage of the experiential learning technology, and the results will be used to define the KPI's displayed on the dashboard.

2.4 Visualization

The dashboard is divided into three cards. The first contains indicators about individual students, the second is about the team, and the third is a visualization based on the regression analysis (See Figure 1). The first shows individual indicators including submissions, assessments and achievements. The second shows team indicators including team submissions, assessments and achievements. The third is based on Industry Sponsor feedback and the regression analysis. partner) regarding the final Virtual Gemba walk.

The students can view their own indicators and their team's. The educator and industry partner can choose which team and student they want to visualize.

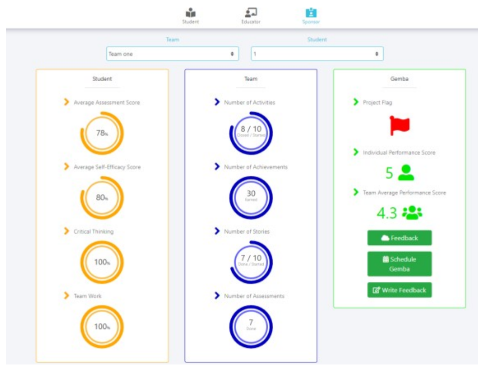


Figure 1: Virtual Gemba Analytics Dashboard for Experiential Learning

3 CONCLUSION

The purpose of the Virtual Gemba Walk is to improve students' success in experiential learning project. With the Virtual Gemba Walk, students will get a more real professional experience as they are presenting to and receiving feedback from the industry partner often. Additionally, the Industry sponsor will have a greater understanding of the project development, giving them more confidence in the success of the project.

4 ACKNOWLEDGEMENT

This work is supported by the National Science Foundation under award DUE-1725941. However, any opinions, findings, conclusions, and/or recommendations are those of the investigators and do not necessarily reflect the views of the Foundation.

5 REFERENCES

Gasevic, D., Dawson, S., & Siemens, G. (2014). Let's not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64-71. doi:10.1007/s11528-014-0822-x

James, N., Humez, A., & Laufenburg P. (2020). Technology to Scaffold Experiential Learning. *TechTrends, 64*, 636-645.

Petruska, R. (2018). *Gemba Walks for Service Excellence*. Productivity Press.

Reynolds, M. (2009). Wild Frontiers—Reflections on Experiential Learning. *Management Learning, 40*(4), 387-392. doi:10.1177/1350507609335848

Scholes, V. (2016). The ethics of using learning analytics to categorize students on risk. *Educational Technology Research and Development, 64*(5), 939-955. doi:10.1007/s11423-016-9458-1

Verner, J., Evanco, W., & Cerpa, N. (2007). State of the practice: An exploratory analysis of schedule estimation and software project success prediction. Information and Software Technology, 49(2), 181-193. doi:10.1016/j.infsof.2006.05.001

Introducing Real-Time Visualization Methods of Learning Support Behaviors for in-Classroom Lessons toward Optimized Assistance

Ryuichiro Imamura, Yuuki Terui, Shin Ueno and Hironori Egi
The University of Electro-Communications
Ryuichi.Imamura@mail.uec.jp, hiro.egi@uec.ac.jp

ABSTRACT: Since each student in a class requires a different degree of learning support, teaching assistants (TAs) need to identify and remember the status of learning support for each student individually. However, it is difficult for TAs to interpret the status of learning support for each individual student, especially in simultaneous lessons. In this study, we developed a system to enable the TAs to figure out the status of their own learning support. The system visualizes their learning support behaviors in real time, and it gives feedback during class time. The results of the practice in the lesson confirmed that the TAs were able to figure out the status of the learning support. Some of the TAs reported that they also changed their learning support behaviors during the process.

Keywords: Teaching Analytics, Teaching Assistant, In-Classroom Lesson, Spatial Usage

1 INTRODUCTION

The teaching assistant (TA) has been widely adopted in universities in order to facilitate effective learning among students. TAs are able to provide more personalized support to students as they work on exercise lessons, such as answering questions or giving hints. The degree of learning support required by a student varies individually. In addition, it is necessary for TAs to pay attention to the assistance dilemma (AD) when providing learning support (K. Miwa et al., 2012). Therefore, TAs need to keep track of the degree of learning support provided to each student during class time, which is especially difficult in the case of simultaneous classes.

The research of teaching analytics is conducted to solve this problem. One of the aims of teaching analytics is to analyze the behavior of teachers, and then provide feedback to them in order to improve their teaching methods. A case study is presented to automatically estimate teacher's behavior (Luis et al., 2016). Their research demonstrated the feasibility of systematically estimating activities by collecting multimodal data sets. There is a study focusing on the behavior of teachers as well as their location in the classroom. A method was proposed to visualize the location history of teachers based on the spatial pedagogy of how they utilize the classroom space during class time (Roberto, 2019). In an experiment conducted in a group work class, there was a gap in the ratio of time spent in contact with each group. The gap existed between the teacher's perception and the results of the system. This indicates that the use of the teacher's positional history may help to change teaching strategies.

The purpose of this study is to enable TAs to figure out the learning support they have provided during class. The proposed system visualizes the TA's learning support behaviors and location history in real time. The in-class practice examined whether the system allows TAs to figure out their own learning support status, and whether any behavioral changes occurred by using the system.

2 METHODOLOGY

Figure 1 shows the overflow of the proposed method. The TA equips a wearable device, shown in Figure 2, during the lesson. The system estimates the learning support behaviors of the TA at the time, from the measured data of the wearable device. The wearable device consists of a tracking camera (Intel RealSense T265) and a Raspberry Pi 4. The learning support behaviors are divided into four states: "Instructing," "Monitoring," "Walking," and "Standby." Instructing refers to the state in which the TA is providing learning support with direct communication to the student. Monitoring refers to a situation where the TA observes the learning situation without communicating. Walking refers to desk-to-table patrol. Standby refers to the state in which the TA is waiting in their seat, not moving around, etc. The behavioral estimation was accomplished using the random forest method, which utilizes posture data compiled from eight graduate students who have experienced being a TA in face-to-face classes of the previous year. Following data captured by the tracking camera were used as explanatory variables; vertical position, three axis velocity, horizontal velocity, pitch, yaw and roll.

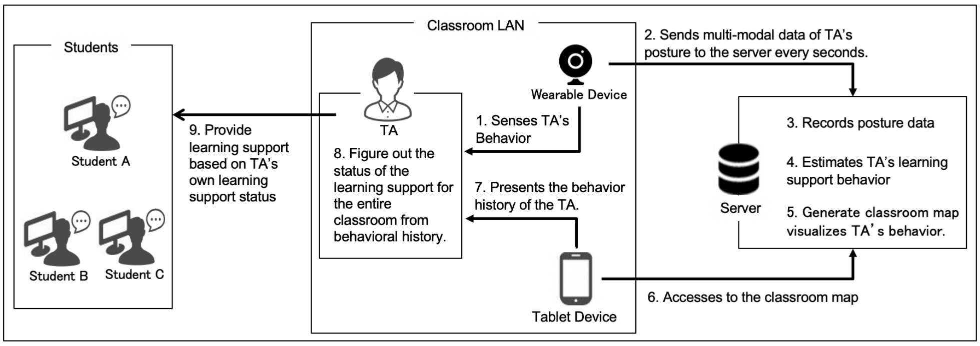


Figure 1: Overview of the method

The TAs receive feedback on their learning support behavior history found on the classroom map shown in Figure 3. On the classroom map, the location history of the TA during class time is plotted by color coding for each learning support behavior that the TA expressed at each location. The TA carries a tablet that shows this map in the lesson, and checks the map to understand their own learning support behavior in real time.

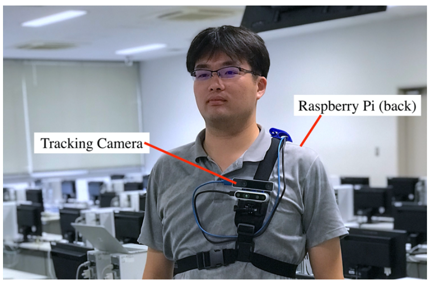


Figure 2: Wearable device

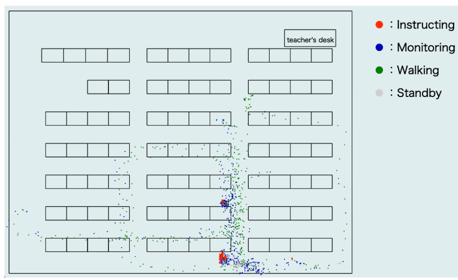


Figure 3: Generated classroom map

3 PRACTICE IN CLASSROOM LESSONS AND RESULTS

The system was introduced in a programming exercise lesson for first-year undergraduate students at a university of science and technology. One teacher and two TAs conducted the lesson, and students' attendance was optional (due to infection control). Seven classes dealing with the same content and 13 TAs were included in the analysis. In the experiment, TAs were only briefed about the system UI, and they provided learning supports based on their own decisions by looking at the map. After the lessons, TAs were interviewed to see if they understood their own learning support situation.

As a result of implementing the system in the lessons, a behavioral visualization like the one in Figure 3 was generated. Findings from the interviews suggested that the TAs were able to use the system to figure out the students they instructed, and the general percentage of time they devoted to each student. Even though the system does not directly indicate the location of the student to whom the TA instructed, the system made it easier for TAs to recall their memories, and figure out which students they instructed. In this experiment, TAs were not instructed on how they should behave after using the system, but we investigated what they think and how they behave by looking at the UI. We also asked TAs about their own policy of educational support. As a result, it was found that the TAs were more likely to take learning support actions based on their own policy while using the system as a hint. To provide AD-sensitive learning support, we expected that it would be necessary to provide support for students who lacked their own, and to interrupt instructions for students who had already been provided adequate instructions. However, most of the TAs analyzed were negative about providing support voluntarily to the students who do not ask TAs for help. As a result, they do not behave in a way that calls on students after checking their learning support status. On the other hand, some TAs took action to make students feel more comfortable about asking questions by walking the classrooms uniformly while also checking the system. This suggests that some kind of learning support was provided by TAs using the system, even for students who were not directly assisting them.

4 LIMITATIONS AND FUTURE WORK

The developed system is intended to contribute to the situation, where TAs would provide learning support to a large number of students. However, the number of students was greatly limited because the experiment was conducted in classes under COVID-19 infection control. Therefore, it was easier than usual classes for TAs to figure out the learning support status for each student. Future research will include long-term, quantitative observations of the influence on decision-making.

ACKNOWLEDGMENTS

This work has been partly supported by the Grants-in-Aid for Scientific Research (NO. JP19H01710) by MEXT (Ministry of Education, Culture, Sports, Science and Technology) in Japan.

REFERENCES

K. Miwa, Terai Hitoshi, and R. Nakaike. (2012). Tradeoff between problem-solving and learning goals: Two experiments for demonstrating assistance dilemma. In Proc. of the Annual Meeting of the Cognitive Science Society, 2008-2013.

Luis P. Prieto, K. Sharma, P. Dillenbourg, and M. Jesus. (2016). Teaching analytics: Towards automatic extraction of orchestration graphs using wearable sensors. In Proc. of LAK16, 148-157.

Roberto Martinez-Maldonado. (2019). “I Spent More Time with that Team”: Making Spatial Pedagogy Visible Using Positioning sensors. In Proc. of LAK19, 21-25.

Student and Faculty Perceptions of Data That Should and Should Not Be Collected at Universities

Rebecca Arlene Thomas
Oregon State University
rebecca.thomas@oregonstate.edu

Marla Wilks
University System of Georgia
mwilks@ecampus.usg.edu

ABSTRACT: This poster presentation summarizes results from a mullti-institutional qualitative project examining stakeholder perceptions of learning analytics in higher education. The current study focused on student and faculty perceptions of data that should and should not be collected at universities. To do this, we analyzed interview responses from 20 students enrolled in three higher education institutions in the United States, as well as 10 faculty employed at seven higher education institutions in the United States. We examined student and faculty responses to four interview questions that asked for perceptions of "learner" data that should and should not be collected, as well as "instructor" data that should and should not be collected. Qualitative data analysis involved coding the interview responses using holistic coding with an attributional layer, and tallying top responses for each stakeholder group. Results suggested that many stakeholders agree that student engagement and satisfaction data should be collected, while perceptions varied surrounding the collection of student demographic information and performance. Additionally, the majority of participants agreed that instructor data measuring teaching performance should be collected. Additional rounds of coding will consider nuances in participant responses, as well as participant commentary given in combination with responses.

Keywords: learning analytics, higher education, data, ethics, students, faculty, qualitative

1 INTRODUCTION

The field of learning analytics aspires to use data to understand the learning process and enhance student learning (Dawson, Joksimovic, Poquet, & Siemens, 2019). Proponents suggest that learning analytics can transform higher education (online, hybrid, and face-to-face) at the student, instructor, and institutional level by providing easily accessible data paired with actionable solutions (Siemens, 2013). However, questions remain regarding how to best use learning analytics in effective and ethical ways (e.g. Slade & Prinsloo, 2013). For example, data privacy can be understood as a three-part relation between a certain domain of data, people who have privacy related to that data, and other people who have access to that data (e.g. Rubel & Jones, 2016). This suggests that data privacy can only be understood in the context of all three components, and in the field of learning analytics, a lack of involvement from data subjects can undermine the trustworthiness of the collection and use of data (Drachsler & Greller, 2016). The current study addresses this area, as we investigated student and faculty perceptions of what data "should" and "should not" be collected at universities. This allowed us to investigate the data subjects' perceptions of domains of data that they think are appropriate to be used by personnel at their institutions.

2 METHODOLOGY

This study was part of a multi-site interview study that investigated stakeholders' perspectives surrounding learning analytics in higher education. For the current study, we analyzed responses from 20 students and 10 faculty that described data that they think should and should not be collected about learners and instructors in higher education. All of the data were collected from March to September 2020, during the COVID-19 pandemic.

We recruited 20 student participants from three higher education institutions located in different areas in the United States. Students were eligible to participate if they were currently enrolled as degree-seeking students with more than one year (2 semesters or 3 quarters, not including summer terms) of experience at the institution. We recruited 10 faculty participants from seven higher education institutions located in different areas of the United States. Faculty were eligible to participate if they were full- or part-time faculty with a minimum of 2 years consecutive teaching experience at the institution (4 semesters or 6 quarters, not including summer terms). Student and faculty participants completed 60 minute interviews via Zoom. For the current study, participant responses to four interview questions were qualitatively coded using holistic coding with an attributional layer (Saldaña, 2016). Table 1 describes relevant codes

Table 1: This table describes relevant codes for "learner data" and "instructor data."

Code	Description	Examples
Demographic information	Relatively stable characteristics about students/instructors used to group individuals	gender, race/ethnicity, SES, age, sexuality, parent demographics
Student/ instructor satisfaction	Feedback about campus experiences, as well as in courses	evaluations, surveys, feedback
Student performance	Data about students' performance in their college coursework	final grades, quiz grades, feedback on course assignments
Teaching performance	Evaluation data about instructors' teaching behaviors and past performance	teaching evaluations, student success, responsiveness
Instructor qualifications	Data related to instructors' professional experience and expertise	educational history, teaching history, degrees
Student engagement	Data about students' behaviors that indicate participation and effort levels	Timeliness, tardiness, attendance, LMS interactions
Educational history	Data about students' academic performance prior to current course enrollment	standardized test scores, past credits, past course failures
Personal life information	Information about students' life circumstances external from the university environment	stress, physical and emotional illness, life events, disability

3 RESULTS

Top codes included in Table 2 were mentioned by at least 15% of at least one stakeholder group.

Table 2: Student and Faculty Perceptions of Data That *Should* and *Should Not* Be Collected

	Learner Data (data about students)		Instructor Data (data about faculty)	
	Should Be Collected	Should Not Be Collected	Should Be Collected	Should Not Be Collected

Code	% of students (N=20)	% of faculty (N=10)	% of students (N=20)	% of faculty (N=10)	% of students (N=20)	% of faculty (N=10)	% of students (N=20)	% of faculty (N=10)
Demographic information		20%	30%	20%			30%	20%
Student satisfaction	50%	60%			40%			
Instructor satisfaction						20%		
Student performance	50%	40%	25%					
Teaching performance	20%	20%			95%	90%		
Instructor qualifications					20%			
Student engagement	30%	30%				20%		
Student educational history				30%				
Personal life information				40%			15%	

4 CONCLUSION

Results reveal varied perceptions of the usage of learning analytics. Many faculty and students mention that student engagement and satisfaction data should be collected while the results are less clear about student demographics and student performance. Respondents also considered student satisfaction, teaching performance, and student engagement data to be both learner and instructor data, suggesting that distinctions between the learner and instructor categories may not be clear cut. The most universal finding related to teaching performance, with 95% of students and 90% of faculty believing that this data should be collected. Additional coding will consider the lack of emergent themes in response to what instructor data should not be collected, the nuances behind the conflicting answers about student performance and demographics, and participant commentary and discussion given in combination with participants' responses. Some of this commentary may suggest burdens and benefits of collecting certain kinds of data, awareness of data collected, as well as specific personnel that they think should have access to that data (e.g. it is possible that they may think certain data is appropriate for advisors to use, but not instructors). Future research can further investigate the impact of collecting different data, combined with perceptions of data sensitivity, as that information may explain why different stakeholders think data "should" or "should not" be collected.

REFERENCES

Dawson, S., Joksimovic, S., Poquet, S. & Siemens, G. (2019). Increasing the impact of learning analytics. In *The 9th International Learning Analytics and Knowledge Conference (LAK19)*, (pp. 1-10). New York, NY USA: ACM.

Drachsler, H., & Greller, W. (2016). 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* (pp. 89-98).

Rubel, A., & Jones, K. M. (2016). Student privacy in learning analytics: An information ethics perspective. *The Information Society, 32*(2), 143-159.

Saldaña, J. (2016). *The coding manual for qualitative researchers.* Sage.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380-1400.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1510-1529.

Extensive Reading Using an E-Book System and Online Forum

Chifumi Nishioka[1,*], Hiroaki Ogata[2]
[1] Kyoto University Library, Kyoto University, Japan
[2] Academic Center for Computing and Media Studies, Kyoto University, Japan
[*]nishioka.chifumi.2c@kyoto-u.ac.jp

ABSTRACT: This paper presents an extensive reading project conducted on an e-book system. We use picture books and comic books as reading materials, and provide an online forum where students can share and discuss their impressions of these. As initial results of the project, we show students' reading patterns, the influence of the online forum on reading amounts, and the influence of reading amounts on performance. The results indicate that the forum may stimulate students and encourage them to continue doing extensive reading. We also observed moderate correlations between the reading amounts and exam scores.

Keywords: extensive reading, e-book, language learning

1 INTRODUCTION

Extensive reading is defined as independent reading of a large quantity of materials for information or pleasure (Renandya et al., 1998)—different studies, such as the one by Nishizawa et al. (2010), have reported that it is effective in the acquisition of a second language. In recent years, digital books (i.e., e-books) have been introduced in schools in different countries. E-book system activities are recorded as learning logs that are used for learning analytics. This paper presents an extensive reading project conducted on an e-book system—we use BookRoll (Ogata et al., 2015). BookRoll is a web application that provides digital learning materials (e.g., textbooks and slides) on students' devices (e.g., tablets and laptops). We use picture books and comic books as reading materials for extensive reading, and students can share their impressions of these on an online forum. As initial results of the project, we show students' reading patterns, the influence of the online forum on reading amounts, and the influence of reading amounts on performance.

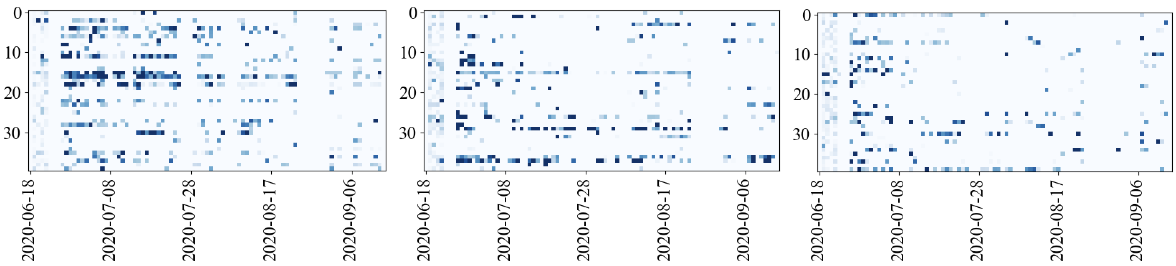
2 THE EXTENSIVE READING PROJECT

The extensive reading project began in June 2020. It focuses on 120 first-year students in three classes at a junior high school in Japan. As reading materials, we provide picture books and comic books, as Hafiz and Tudor (1989) reported that shorter books place less strain on learners' concentration, and are thus more likely to be preferred. Before starting the project, the lecturer showed a short movie on the principles of extensive reading (e.g., "learners choose what they want to read") (Day & Bamford, 1998), and as a starter, recommended a series of comic books. The project provides an online forum where students are encouraged to share and discuss their impressions of the reading materials. All three classes are provided the same reading materials and instructions.

3 INITIAL RESULTS

3.1 Reading Patterns

Figure 1 shows students' reading patterns for each class. The x-axis and y-axis correspond to the date and student, respectively. Each cell represents the number of pages read by a student on the given date, with dark colors indicating a larger number of pages. We can see that students in Class A read the materials more frequently and extensively than those in the other classes.

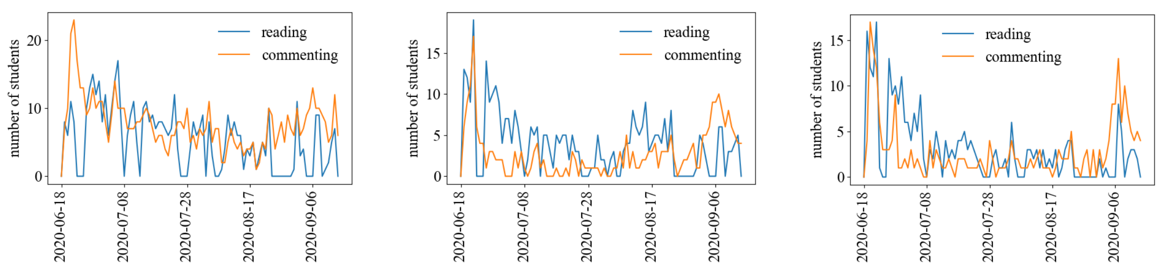


(a) Class A (b) Class B (c) Class C

Figure 1: Reading patterns

3.2 Influence of the Online Forum on Reading Amounts

As described in Section 2, the project provides students with an online forum to share and discuss their impressions of the reading materials. Figure 2 presents the numbers of students who did extensive reading (blue line) and who posted a comment on the forum (orange line) on the given date. We can see weak or moderate correlations between them (r=0.26, p=.01 for Class A; r=0.33, p<.01 for Class B; r=0.45, p<.01 for Class C). Figure 2(a) shows that the number of students posting a comment is continuously high for Class A. According to the lecturer, Class A students have shown greater inclination to continue doing extensive reading, indicating that the forum may stimulate the students and encourage them to continue doing extensive reading.

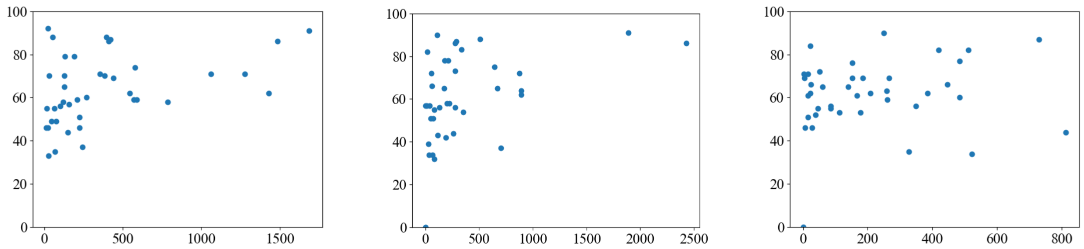


(a) Class A (b) Class B (c) Class C

Figure 2: The numbers of students who did extensive reading and who posted a comment on the online forum

3.3 Influence of Reading Amounts on Performance

We analyzed the influence of reading amounts on the students' language learning performance—an English exam was conducted on September 30, 2020; its content was not related to that of the reading materials. Figure 3 shows the correlations between reading amounts and the exam scores. Each plot in Figure 3 corresponds to each student. The x-axis and y-axis indicate the number of pages read by a student and the exam score, respectively. In addition, we calculated the Pearson correlation coefficients that demonstrated moderate positive correlations (r=0.40, p=.01 for Class A; r=0.41, p=.01 for Class B), with Class C (r=0.14, p=.38) being the exception. Figure 3 shows the students who have low exam scores even though they read many pages. In the future, we would like to investigate how to improve these students' learning, and how to motivate those who did not extensive read and received low scores.



(a) Class A (b) Class B (c) Class C

Figure 3: Scatter plot showing reading amounts (x-axis) and exam scores (y-axis)

4 CONCLUSION

This paper presents the initial results of an extensive reading project on an e-book system. While students do extensive reading independently, sharing their impressions of the reading materials with their peers may contribute to promoting extensive reading. We will continue with the extensive reading project and reveal factors that contribute to the acquisition of a second language.

ACKNOWLEDGEMENTS

This work was supported by the Inamori Foundation, NEDO SIP 18102059-0, JSPS KAKENHI Grant Numbers 16H06304 and 18K13235, and JST AIP Grant Number JPMJCR19U1.

REFERENCES

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

Nishizawa, H., Yoshioka, T., & Fukada, M. (2010). The impact of a 4-year extensive reading program. In JALT2009 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. *RELC Journal, 29*(2), 187–191. https://doi.org/10.1177/003368829802900211

Student Response Estimation using E-book Reading Logs with Textbook Information

Tsubasa Minematsu*, Atsushi Shimada, and Rin-ichiro Taniguchi
Kyushu University, Japan
*minematsu@ait.kyushu-u.ac.jp

ABSTRACT: We estimated student responses about their comprehension of the digital textbook based on e-book reading logs. Knowing their comprehension helps to provide various educational supports for both students and teachers. However, it is difficult to estimate them accurately because of insufficient reading logs. In this study, we investigated the effectiveness of additional information for the estimation of their understanding. Image data of pages and text data in textbooks were used as the additional information and were combined with reading logs using a deep neural network. In this experiment, we confirmed the effectiveness of the combination of these data and discussed the potential of reading logs for knowing student understanding.

Keywords: reading log, comprehension analysis, multimodal data, neural network

1 INTRODUCTION

Understanding student's comprehension of learning contents can help support them and improve teaching materials that students find difficult. Recently, e-learning systems collect various learning logs as the learning activities of students. For example, BookRoll (Ogata et al., 2015) directly collects student's responses on each page. Students can push the 'getit' or 'notgetit' button provided by BookRoll depending on their understandings. However, students do not always use the response functions. It is difficult to grasp students' understanding of pages with a few responses. The other approach is to estimate student's comprehension based on e-book reading logs representing student-system interactions. However, it is difficult to estimate comprehension because of the quality of the reading logs. The reading logs do not always represent the cognitive process of students. Even if two students' reading logs are similar in different contents, they may have different comprehensions. Feng, D'Mello, & Graesser (2013) reported that "mind wandering interacted with text difficulty in predicting reading comprehension". The only reading logs-based method is limited, and we need to combine additional information. In this study, the effectiveness of additional information is investigated. Texts and images of textbooks are chosen as additional information because we focus on understanding student's comprehension in self-study before classes. The textbook information is an important factor in the self-study. We propose a deep neural network for combining the reading logs and the textbook information. The accuracy of our combination method is evaluated on each page and each student's reading log.

2 DEEP NEURAL NETWORK USING READING LOGS AND TEXTBOOKS

Reading logs and ' getit/notgetit' responses were collected from 95 students who took a seven-week course in cybersecurity as first-year students in our university's school of design. The students read a digital textbook provided by BookRoll, and they could add highlights and notes on each page. The reading logs and their responses were stored in the BookRoll database. The reading log for each page was used during self-study before class except the first class because it was class guidance. The reading log on each page represents the number of operations by the students such as 'next page' and 'add highlight' and reading time. In addition, we used images and text on each page from the textbook. As the text information, we counted the number of words that appeared on each page to make a one-hot vector. As the image data, we capture the pages of the textbook. In this study, we were able to collect 9589 reading logs for 'getit' and 954 for 'notgetit'. The number of pages was 364.

We used a deep neural network, as shown in Figure 1, because it is easy to combine heterogeneous data, such as reading logs and image data. The neural network receives the reading log vector, the image, and the text vector of the page. Our network extracts a feature vector of each data on each path and integrates the extracted features for estimating the probability of whether the student responded to 'getit' or 'notgetit' on each page. In this study, 70% of the collected data were used as training dataset and the rest as test dataset. We chose the data randomly when making the datasets. The training dataset and the test dataset may contain different data from the same student. In the training phase, our network learned relationships between reading logs, textbook information, and responses from students. Note that we adopted an oversampling strategy because the number of 'getit/notgetit' were imbalanced. In our training phase, a mini batch contained the data of 10 'getit' responses and 10 'notgetit' responses.

3 RESULTS AND DISCUSSION

To investigate the effectiveness of additional information combinations, we evaluated the proposed method using the test dataset. First, we conducted a page-wise evaluation. Each page has different distributions of the response 'getit' and 'notgetit'. If our network can estimate the distribution from the reading logs and the textbook information, we can understand whether the page is difficult or not. For the page-wise evaluation, we estimated the probability of the response in the test dataset and averaged the probability on each page. The averaged probability was compared to the distribution computed from the actual observed response in the test dataset. We used Kullback–Leibler divergence (KLD) for the comparison. As an ablation study, we compared our proposed network (rib-net) to r-net with reading logs only and ib-net with textbook information only. Figure 2 shows the KLD distribution approximated by a Gaussian distribution. The KLD value means that the closer to zero, the more accurate the distribution is estimated. Rib-net captured the trend more correctly than ib-net and r-net. We observed that the probability computed by ib-net was biased towards either response. Second, we conducted a reading log-wise evaluation. If we can estimate a student's response using the reading log and textbook information, we can identify the student and pages (contents) the student finds difficult. Since the output value of our network has two probability values of 'getit' and 'notgetit', we need to decide whether to 'getit' or 'notgetit'. In this study, we selected the one with the highest probability in 'getit' and 'notgetit'. To compute the chance rates, we used the classifier which outputs 'getit' or 'notgetit' with a 50% possibility. Precision and Recall, F-measure were used as evaluation metrics. The result is shown in Table 1. Rib-net exhibited higher results than the chance rate, however, ib-net had better performance than rib-net. Note that this result does not imply that ib-net was better because ib-net could not provide appropriate estimation as shown in Figure 2.

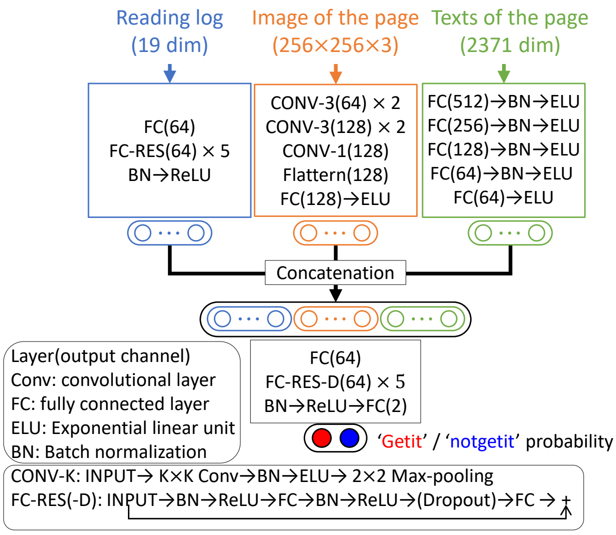


Figure 1: Network architecture

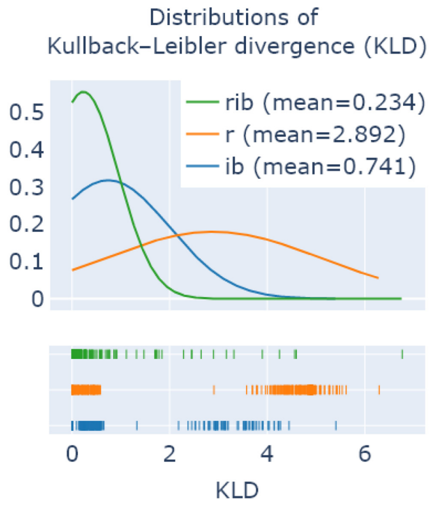


Figure 2: The page-wise evaluation

Table 1: Comparison of the reading log-wise evaluation.

	P-G	R-G	F-G	P-N	R-N	F-N
chance rate	0.902	0.500	0.643	0.099	0.501	0.165
r-net	0.905	**0.916**	**0.911**	0.139	0.123	0.131
ib-net	**0.973**	0.798	0.877	**0.302**	**0.798**	**0.438**
rib-net	*0.948*	*0.850*	*0.896*	*0.295*	*0.572*	*0.389*

*Bold and italic scores mean the best and the 2nd best. Precision (P), recall (R), F-measure (F), getit (G), and notgetit (N).

These results indicated that the combination of the reading logs and the textbook information was an effective way for predicting students' responses. In this experiment, the students' responses of each page could be grasped by using rib-net, which may be able to help teachers to improve their textbooks. In the reading log-wise evaluation, the combination method contributed to improving the estimation for each student on each page, however, the accuracy was not sufficient. We considered that the performance of rib-net was limited by the presence of data with different responses for similar reading logs. We observed that most students mainly read the textbook without any other actions. It is difficult to distinguish between such data. In the comprehension estimation, the quality of current reading logs was insufficient. A solution would be to use the other additional data such as memo texts and reflections. Another solution is a data collection mechanism that encourages active learning behavior. These solutions are aimed at obtaining distinguishable learning logs for the analysis.

ACKNOWLEDGEMENTS

This work was supported by JSPS KAKENHI Grant Number JP19K20421.

REFERENCES

Feng, S., D'Mello, S. & Graesser, A.C. (2013). Mind wandering while reading easy and difficult texts. *Psychonomic Bulletin & Review* 20, 586–592. https://doi.org/10.3758/s13423-012-0367-y

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.

Providing Personalized Nudges for Improving Comments Quality in Active Video Watching

Negar Mohammadhassan
University of Canterbury
Negar.mohammadhassan@pg.canterbury.ac.nz

Antonija Mitrovic
University of Canterbury
Tanja.Mitrovic@canterbury.ac.nz

ABSTRACT: This interactive demo presents AVW-Space, an online video-based learning platform which supports engagement by providing personalized interventions called nudges (Mitrovic et al., 2019). AVW-Space has a note-taking area which allows students to write comments on educational videos. Previous studies on AVW-Space showed that a lot of comments made by students are of low quality and merely repeat video content (Mitrovic et al., 2017). The new version of AVW-Space provides nudges to guide students towards writing high-quality comments and self-reflections step-by-step. The nudge framework includes a machine learning model to predict the quality of the comment the student has made (Mohammadhassan et al., 2020). Then a suitable nudge is triggered for the student adaptively, based on the student profile, the history of the nudges the student has received, the timestamp of the video that the student is watching and the history of comments the student has made on the current video. We investigated the effectiveness of these nudges in a study where AVW-Space was used for training on presentation skills. This study showed that the quality nudges improve engagement and learning in students. However, to investigate the effectiveness of these nudges in different domains, the quality nudge framework should be generalized to other fields in future work.

Keywords: Personalized Intervention, Video-based Learning, Natural Language Processing

1 DEMO VIDEO

https://youtu.be/BG1ysuuZFwM

REFERENCES

Mitrovic, A., Dimitrova, V., Lau, L., Weerasinghe, A., & Mathews, M. (2017). Supporting Constructive Video-Based Learning: Requirements Elicitation from Exploratory Studies. *Proc. 18th Int. Conf. Artificial Intelligence in Education* (pp. 224–237). Springer International Publishing.

Mitrovic, A., Gordon, M., Piotrkowicz, A., & Dimitrova, V. (2019). Investigating the Effect of Adding Nudges to Increase Engagement in Active Video Watching. *Proc. 20th Int. Conf. Artificial Intelligence in Education* (pp. 320-332). Springer International Publishing.

Mohammadhassan, N., Mitrovic, A., Neshatian, K., Dunn, J. (2020) Automatic quality assessment of comments in active video watching using machine learning techniques. *Proc. 28th Int. Conf. Computers in Education* (pp. 1-10). Asia-Pacific Society for Computers in Education.

SEET: A Visual Learning Analytics tool for Supporting Equity in Science Classrooms

Ali Raza[1,2], William R. Penuel[2], Tamara Sumner[1,2]
Department of Computer Science[1], Institute of Cognitive Science[2]
{a.raza, william.penuel, sumner}@colorado.edu

ABSTRACT: Experience is an important dimension of learning (Roth & Jornet, 2014). As our classrooms become more diverse and multicultural, more effort is required to ensure instruction is culturally relevant and sustainable for all students. Understanding students' experiences and identities can make classrooms more meaningful for all students irrespective of gender, race, and socio-economic status. We developed a Visual Learning Analytics (VLA) system - the Student Electronic Exit Ticket (SEET) for understanding student experience of the classroom, based on the construct coherence, relevance, and contribution. Which prior research have shown to be a reliable indicator of the equitable student experience (Reiser et al., 2017, NASEM, 2018, Miller et al., 2018). SEET captures and visualizes student experience data revolving around questions based on the above mentioned constructs. It disaggregates data across gender, race and also provides over time tracking abilities with complete anonymization to teachers and researchers. SEET consists of six different data visualizations developed following a co-design partnership with four middle school science teachers and two instructional coaches. We share a demo of our tool currently deployed in six middle school science classrooms for understanding student experience to create equitable and just learning environments.

Link to video: https://drive.google.com/drive/folders/1pFE-irih6pVule-JtDPSfCaJI7gTyA_Z?usp=sharing

Keywords: Equity, Visual Learning Analytics, Design, Student Experience

REFERENCES

Miller, E., Manz, E., Russ, R., Stroupe, D., & Berland, L. (2018). Addressing the epistemic elephant in the room: Epistemic agency and the next generation science standards. Journal of Research in Science Teaching, 55(7), 1053-1075.

National Academies of Sciences, Engineering, and Medicine. (2018). How people learn II: Learners, contexts, and cultures. National Academies Press.

Roth, W. M., & Jornet, A. (2014). Toward a theory of experience. Science Education, 98 (1), 106-126.

Reiser, B. J., Novak, M., & McGill, T. A. (2017). Coherence from the students' perspective: Why the vision of the framework for K-12 science requires more than simply "combining" three dimensions of science learning. In Board on Science Education Workshop "Instructional Materials for the Next Generation Science Standards.

Learning Analytics Dashboard for Monitoring Students' Free-practice Learning Activity

Han Zhang
University of Pittsburgh
haz97@pitt.edu

Jordan Barria-Pineda
University of Pittsburgh
jab464@pitt.edu

Peter Brusilovsky
University of Pittsburgh
peterb@pitt.edu

ABSTRACT: Currently, hybrid instruction models are being adopted by many universities, which have led to the generation of different types of learning data. However, it is hard for instructors to analyze their students' learning given the large data amount. In this demo, we present the initial prototype of an interactive dashboard system where instructors can check students' progress in free-practice learning activities to quickly estimate their learning status. We followed a user-centered design approach for determining the features that a group of programming instructors prioritized to have in a learning analytics dashboard. We delivered a survey prepared based on existing learning analytics literature. Based on the collected requirements, we designed a prototype that aggregates the real-time information at two levels of granularity: per week and per topic. To show how multi-variables affect learning performances jointly, the dashboard uses size and color in a dot matrix to show different learning statuses. A radar graph is adopted to display multivariate data so that instructors can understand the weaknesses and strengths of specific students and the whole class on average. We tested the functionalities of this prototype with two instructors, by using data from a 13-week programming college class taken by 55 undergrad students.

Keywords: Learning analytics, learning dashboards, interactive system design

Video: https://www.youtube.com/watch?v=bZ22aCFPio8&feature=youtu.be

Through the eyes of cooperation at multi-touch tabletop displays

Matthias Ehlenz, Birte Heinemann, Rabea de Groot, Damin Wito Kühn, Claudio Nadenau, Domenic Ulrich Quirl, Ulrik Schroeder
RWTH Aachen University
Ehlenz|Heinemann|Schroeder@cs.rwth-aachen.de

ABSTRACT: Learning Analytics in collocated collaboration situations is challenging in itself. Combined with multi-touch input devices, which do not discriminate among individual users, opens a whole new bundle of challenges. To capture data beyond videography and touch interaction, this project presents the integration of eye-tracking glasses into an open-source framework for multi-touch learning applications. Not only does this help to aggregate data from multiple sources for downstream multi-modal learning analytics, but it also provides (near) real-time access to gaze data in the application itself, opening new opportunities for open-source learning environments reacting to the individual learner in group setups. This data helps to gain insights on the learner-to-learner interaction and helps researchers to understand what learners do while they don't interact with the system, i.e. fixating the last position of interaction as an indicator for reflection of the own behaviour, sweeping gaze in search for the next opportunity for interaction or off-screen gaze likely to be an advice-seeking gaze at the fellow students. It also helps advising single possibly timid learners, when no one else is looking. In this demo, we present the system and are looking forward to discussing further ideas and application scenarios.

Keywords: Eye Tracking, Multi-Touch Games, Collaboration, Interaction, Learning, Serious Games, Large Scale Displays

BACKGROUND AND FURTHER INFORMATION

The video is available at: http://elearn.rwth-aachen.de/RFC2020. For further information about the idea, hardware setup and research background, see (Heinemann et al., 2020). For the framework and research context, see (Leonhardt et al., 2019). For the game, see (Ehlenz et al., 2018).

REFERENCES

Ehlenz, M., Röpke, R., & Schroeder, U. (2018). Kollaboratives Multitouch Lernspiel für Reguläre Ausdrücke. In D. Krömker & U. Schroeder (Hrsg.), *DeLFI 2018—Die 16. E-Learning Fachtagung Informatik* (S. 307–308). Gesellschaft für Informatik e.V.

Heinemann, B., Ehlenz, M., & Schroeder, Prof. Dr. U. (2020). Eye-Tracking in Educational Multi-Touch Games: Design-Based (interaction) research and great visions. *ACM Symposium on Eye Tracking Research and Applications*, 1–5. https://doi.org/10.1145/3379156.3391838

Leonhardt, T., Ehlenz, M., Röpke, R., Bergner, N., & Schroeder, U. (2019). MTLG – Helping students and teachers creating and analysing simulations and games from a didactical and technical perspective. In N. Pinkwart & J. Konert (Hrsg.), *DELFI 2019 – Die 17. Fachtagung Bildungstechnologien* (S. 319–320). Gesellschaft für Informatik e.V. https://doi.org/10.18420/delfi2019_331

Evaluating the acquisition of 21st century skills within online educational settings: A data-driven psychometric approach

Abhinava Barthakur
University of South Australia
abhinava.barthakur@mymail.unisa.edu.au

ABSTRACT: In the digital age, the changing economy demands graduates to master some broad transferable skills along with acquiring subject knowledge, for workplace readiness. As such, there has been increasing calls for teaching important skills (such as critical thinking and leadership) and developing effective means for their assessment. While online learning has had considerable success in enabling lifelong learning and promoting 21st century skills, there is still a significant challenge in the way these skills are assessed. Although assessment plays a critical role in evaluating the accomplishments and systematic learning progression, and has been an integral part of online learning, it has primarily focused on the content-related knowledge, with significant gaps in the assessment of 21st century skills of learners. Furthermore, the pervasive use of educational technology has allowed researchers to collect enormous amounts of learners' performance and process data, thus increasing the possibilities of evaluating the acquisition of 21st century skills. In this regard, we propose a robust framework marrying theory-driven psychometric models and data-driven learning analytics algorithms to assess the proficiency of learning and development of leadership and critical thinking skills within the MOOC context.

Keywords – Learning analytics, psychometric measurement models, 21st century skills, online learning, performance and process data

1 INTRODUCTION

The changing nature of the modern workplace and the recent technological advances have highlighted the need for university graduates to acquire soft-transferable skills. Being well aware that school success is not the only influential factors determining the success of the economic (Kyllonen, 2012), business leaders, employers and educational researchers have called for policies that would support the development of more broad, transferable skills – commonly referred to as the 21st century skills (Vockley, 2007). Skills such as critical thinking, problem-solving, leadership, communication, collaboration are some of the most desirable competencies for future graduates (Casner-Lotto & Barrington 2006; Lai & Viering 2012). The Partnership for 21t Learning (P21) report states that 92.1% and 81.8% of employers consider critical thinking and leadership skills vital for workplace readiness respectively (Casner-Lotto & Barrington, 2006). Critical thinking allows creative and constructive solutions through well-supported claims and evidence, while congruous leadership contribute towards a positive work climate and job satisfaction among employees. Allowing learners to thrive through life's challenges, these skills contribute significantly to their future career prospects and success.

With recent technological advances, online learning has been increasingly seen as a prominent approach in delivering workforce professional development programs (Bond, 2013). In that sense, Massive Open Online Courses (MOOCs) received considerable attention as one of the most prominent modalities for delivering professional development programs. Besides allowing mastery of subject content, MOOCs through their varying pedagogy and self-regulated learning provides learners with tremendous opportunities for developing soft-transferable skills and life-long learning (Chauhan, 2014). The underlying impact of MOOCs in nurturing these highly valued skills in the labour market allows learners to cultivate knowledge and skills beyond a specific domain. For that matter, cooperate learning are shifting their focus towards such modes of learning for developing the necessary workforce skills and professional development within their employees (Bond, 2013).

Ellis (2013), among others, have highlighted that face-to-face mode of course delivery (and its inability to capture learners' interaction data) is a key limitation in improving assessments. With the recent technological advances within educational contexts, have however, made possible to capture both learners' performance and process data. The effective use of these data will provide opportunities to understand the learning progression of learners and their acquisition of 21st century skills.

Although a lot of work has been done on the assessment of those soft-transferable skills, evaluation of learning and development of these skills has however been of increasing concern. Assessments for learning are designed to assess a learner's knowledge about subject content. Casner-Lotto and Barrington, (2006) using their P21 framework have emphasized the need for assessing learning and acquisition of 21st century skills to provide formative intervention to steer and support students' learning. Given that measuring these skills are inherently difficult (Knight et al., 2013), assessment of learning of 21st century skills can be better explained by exposing learners to these simulation-based online learning environments comprising of well-designed tasks, interactions with other peers, and instructors as well with various course components (Rupp et al., 2010).

One particular challenge with the assessment of 21st century skills such as critical thinking and leadership is the lack of coherent understanding of the nature and development of these skills among learners (Care et al., 2018). Theoretical frameworks developed for assessing these skills do not provide a clear understanding of progression through the different stages of their development and how learners' learning is associated with the development. However, there lies a great scope and promise in bringing multi-disciplinary techniques for detecting and measuring 21st century skills within MOOCs. This doctoral research focuses on building robust blended psychometric models that go beyond ranking learners on a continuous scale. The proposed approach draws on data-driven techniques rooted in learning analytics (LA), which – coupled with educational assessment techniques – provide holistic ways to assess the development of critical thinking and leadership skills in MOOCs. Given the value and importance of these skills for workplace readiness, this research will utilize the rich longitudinal data generated within the MOOC platforms to determine the mastery of multiple fine-grained subskills within a broader domain that extends beyond traditional assessment methods of reporting a single score value of learning proficiency. The development of these skills will be further assessed across multiple courses within a study program to evaluate the progression of the learning competencies over time.

2 BACKGROUND

2.1 21st century skills

Unlike classroom activities which are often constrained to learning within disciplinary boundaries, real-world challenges in the workplace typically demand both collaborative and individual learning approaches. According to Dawson and Siemens (2014), 21st century skills are an absolute necessity for workplace success and "key to individual and community wealth and wellbeing within a society" (p. 285). Besides cognitive abilities, there has been a growing proliferation for skills, which are considered necessary for workforce outcomes and other aspects of personal and professional wellbeing (Pellegrino & Hilton, 2012). From the perspective of this research, critical thinking and leadership skills are of particular interest. Critical thinking is a purposeful, self-regulatory judgement process requiring individuals to analyze arguments, draw inferences using reasoning, and evaluating and solving problems (Ahuna & Kiener, 2014). Likewise, leadership is a dynamic and complex enigmatic process providing a sense of cohesiveness and a healthy mechanism for innovation and creativity along with an overarching sense of vision (van Wart, 2003). Since leadership occurs within groups and societies, many studies have relied upon self-reports and peer reviews to measure leadership skills and what constitutes an adequate leader (Rosch et al., 2014). Similarly, assessment of critical thinking among learners and professionals have mostly relied on measurement instruments (Allen et al., 2004; Bissell & Lemons, 2006; Cisneros, 2009). Most of these studies measure the presence (to what degree) or absence of these skills among individuals without considering the learning interactions and processes. The pervasiveness of online learning, on the other hand, provides multiple opportunities to capture both learning interaction and performance data of learners to assess the development of these skills among individuals.

2.2 Learning analytics and learning assessment

From the psychometrics and measurement science perspective, assessment means evaluating what leaners know and have learned through the analysis of their responses to a fixed set of test items. Assessment within LA, on the other hand, implies evaluation of real-time learner behavior while learning and in the learning environment, with the intent of positively impacting the learning processes (Drachsler & Goldhammer, 2020). Although both disciplines have similar goals, they are different in their theoretical and methodological assumptions (Mislevy et al. 2012). Psychometrics follows a top-down approach, starting with defining the targeted skills and attribute, then collecting the observable indicators eliciting these skills are identified and finishing with designing tasks to obtain data for these observable indicators. In contrast, LA follows a bottom-up methodology, where richer data about learning is collected, analyzed and finally, inferences are drawn about the learning processes (Drachsler & Goldhammer, 2020).

With the changing environment within the education sector, many researchers have argued the importance of the intersection of various disciplines to make inferences about a learner's learning progression (e.g., Wilson & Scalise 2016). Learning within online platforms can provide researchers with tremendous opportunities to investigate learning processes with fine-grained resolution (Drachsler & Goldhammer, 2020). While cognitive sciences have enhanced our understanding of what learners' learn and how they develop their skills (Mislevy et al., 2003), assessments can benefit from the technological advances to link observable behaviors (data from logs, discourse, and social interactions) to inferences about learners' learning (Mislevy et al. 2012). Milligan and Griffin (2016), among others, adopted a multidisciplinary approach to examine the quality and effectiveness of learning in MOOCs. The authors operationalized a 21st century learning competency by building a partial credit model using learners' log data. This study shows the potential of intersecting LA and educational assessment to make inferences about the learners' developmental progression of skills within MOOCs. Another important study bringing together educational assessment and LA is the study by Hu and colleagues (2017). Here, the authors adopted a three-step methodology based on an evidence-centered design framework for the assessment of online problem solving among primary school students. This study, although not within the MOOC domain, is a classic example of using both the process and assessment data to evaluate the general problem-solving capabilities of learners.

3 RESEARCH QUESTIONS

Based on the above discussion and the need for assessing the learning and development of 21st century skills within the MOOCs context using a theory-driven psychometric model and data-driven computational models used within LA, this doctoral research aims at answering the following research questions.

RQ1*: How can the learning of different critical thinking and leadership skills within MOOCs be measured?*

RQ2*: Which analytical features are best indicative of the development of critical thinking and leadership skills, and can this development be traced?*

RQ3*: How can a learner's progression across levels of proficiency be observed within a MOOC?*

4 METHODOLOGY

This doctoral research will use the data from two existing MOOC programs consisting of four courses each designed to address the professional development needs of a US global organization with worldwide presence. The programs focused on supporting professionals in developing leadership and critical thinking skills. Weekly activities within courses were designed to teach certain learning objectives within the broader domain of the two abovementioned 21st century skills. Course designers developed two different kinds of mappings – the first mapping links the assessment items to the course learning objectives (CLO), while the second mapping links CLO to the program learning objectives (PLO). The mappings at the program-level will allow us to investigate the learning behavior and proficiency within a single course and across four courses within a study program.

Our first research question aims at evaluating the levels of learning proficiency learners have developed while studying 21st century skills in a MOOC setting. To answer this question, we adopt a cognitive diagnostic model (CDM) to classify the learners based on their mastery of the CLOs and PLOs. Analyzing the learning proficiency at a course and program level will provide more opportunities to make detailed inferences about their learning progression and how these patterns differ while mastering leadership and critical thinking abilities.

To address RQ2, fine-grained features will be extracted from the trace data that will act as evidence of the development of the two skills while learning within MOOC settings. Mappings will be developed and CDM will be applied to determine the developmental progression of these two skills. By the end of this study, based on the analysis and output, we aim to propose a generalized methodology that can be used to detect the development of leadership and critical thinking skills among learners in any context. A pipeline will be outlined that can be used by other studies to measure the development of these two skills based on the engagement patterns of learners within an online course.

Utilizing both the process and assessment data, RQ3 aims at mapping the results from the first two questions. That is, analyze the change in levels of learning proficiency with the changing learning behavior as indicated by the extracted analytical features. Based on a learner's current state of proficiency and their level of engagement, we aim to develop a robust methodological framework that would inform learners their current level of understandings and skill development, and how further engagement with various course components will help them attain higher levels of proficiency in critical thinking and leadership skills.

5 CURRENT PROGRESS

As a part of answering RQ1, we are currently working on the validation of the mappings, designed by the course providers, linking the course components to the different course learning objectives. The correct specification of these mappings is very crucial as inaccurate mappings can severely alter the parameter estimation and diagnostic classification of learners. The process of validating the mappings will also help the instructors and course providers to enhance the course/program design in their future offerings.

6 CONTRIBUTION

The primary purpose of this doctoral research is centered around utilizing approaches from various disciplines (learning assessment, cognitive science and LA) to develop methods to support the assessment of learning and development of complex 21st century skills within an online learning context.

Theoretical Contribution: While there is a prominent call within online learning (and MOOCs) to explore the development and learning of 21st century skills, there is lack of empirical evidence that studies how online learning enhances the acquisition of leadership and critical thinking skills among professionals. Positioning educational assessment as a part of the learning experience, this doctoral research contributes to the conceptualization of designing assessments using both process and performance data. The theory-driven conceptualizations of the skill constructs will provide a means of understanding the different levels of proficiency among individuals and within groups.

Methodological Contribution: Central to this research's methodological contributions in the field of LA is the work of developing a robust framework for the assessment of learning proficiency within the online, digital educational settings, adopting a blended assessment methodology, comprising of measurement sciences and computational models. The use of analytical features extracted from the learners' trace data to operationalize the 21st century skill constructs will further help describe the progression of learning proficiency on the likely engagement and mastery of different skills. Empirical validation of the proposed framework will allow researchers to use a similar methodology to operationalize 21st century skills within another learning context.

7 REFERENCE

Ahuna, K. H., & Kiener, M. (2014). *A New Era of Critical Thinking in Professional Programs*. *7*(3), 9.

Allen, G. D., Rubenfeld, M. G., & Scheffer, B. K. (2004). Reliability of assessment of critical thinking. *Journal of Professional Nursing*, *20*(1), 15–22.

Bissell, A. N., & Lemons, P. P. (2006). A New Method for Assessing Critical Thinking in the Classroom. *BioScience*, *56*(1), 66–72. https://doi.org/10.1641/0006-3568(2006)056[0066:ANMFAC]2.0.CO;2

Care, E., Griffin, P., & Wilson, M. (Eds.). (2018). *Assessment and Teaching of 21st Century Skills: Research and Applications*. Springer International Publishing. https://doi.org/10.1007/978-3-319-65368-6

Casner-Lotto, J., & Barrington, L. (2006). Are They Really Ready to Work? Employers' Perspectives on the Basic Knowledge and Applied Skills of New Entrants to the 21st Century U.S. Workforce. In *Partnership for 21st Century Skills*. Partnership for 21st Century Skills.

Chauhan, A. (2014). *Massive Open Online Courses (MOOCS): Emerging Trends in Assessment and Accreditation*. 12.

Cisneros, R. M. (2009). Assessment of Critical Thinking in Pharmacy Students. *American Journal of Pharmaceutical Education*, *73*(4). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2720362/

Dawson, S., & Siemens, G. (2014). Analytics to Literacies: The Development of a Learning Analytics Framework for Multiliteracies Assessment. *International Review of Research in Open and Distance Learning*, *15*, 284–305. https://doi.org/10.19173/irrodl.v15i4.1878

Drachsler, H., & Goldhammer, F. (2020). Learning Analytics and eAssessment—Towards Computational Psychometrics by Combining Psychometrics with Learning Analytics. In D. Burgos (Ed.), *Radical Solutions and Learning Analytics: Personalised Learning and Teaching Through Big Data* (pp. 67–80). Springer. https://doi.org/10.1007/978-981-15-4526-9_5

Ellis, C. (2013). Broadening the scope and increasing usefulness of learning analytics: The Case for assessment analytics. *British Journal of Educational Technology, 44*(4), 662-664.

Hu, Y., Wu, B., & Gu, X. (2017). Learning analysis of K-12 students' online problem solving: A three-stage assessment approach. *Interactive Learning Environments*, *25*(2), 262–279.

Knight, S., Buckingham Shum, S., & Littleton, K. (2013). Epistemology, pedagogy, assessment and learning analytics. *Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK '13*, 75.

Kyllonen, P. C. (2012). Measurement of 21st century skills within the Common Core State Standards. *In Invitational Research Symposium on Technology Enhanced Assessments*, 24.

Lai, E. R., & Viering, M. (2012). Assessing 21st Century Skills: Integrating Research Findings. In *Pearson*. Pearson.

Milligan, S. K., & Griffin, P. (2016). Understanding Learning and Learning Design in MOOCs: A Measurement-Based Interpretation. *Journal of Learning Analytics*, *3*(2), 88–115. https://doi.org/10.18608/jla.2016.32.5

Mislevy, R. J., Behrens, J. T., Dicerbo, K. E., & Levy, R. (2012). Design and Discovery in Educational Assessment: Evidence-Centered Design, Psychometrics, and Educational Data Mining. *Journal of Educational Data Mining*, *4*(1), 11–48.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). Focus Article: On the Structure of Educational Assessments. *Measurement: Interdisciplinary Research & Perspective*, *1*(1), 3–62.

Pellegrino, J. W, & Hilton, M. L. (2012). *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century* (p. 13398). National Academies Press.

Rosch, D. M., Collier, D. A., & Zehr, S. M. (2014). Self-vs.-Teammate Assessment of Leadership Competence: The Effects of Gender, Leadership Self-Efficacy, and Motivation to Lead. *Journal of Leadership Education*, *13*(2), 96–124.

Rupp, A. A., Gushta, M., Mislevy, R. J., & Shaffer, D. W. (2010). Evidence-centered Design of Epistemic Games: Measurement Principles for Complex Learning Environments. *The Journal of Technology, Learning and Assessment*, *8*(4), Article 4.

van Wart, M. (2003). Public-Sector Leadership Theory: An Assessment. *Public Administration Review*, *63*(2), 214–228. JSTOR.

Vockley, M. (2007). Maximizing the Impact: The Pivotal Role of Technology in a 21st Century Education System. In *Partnership for 21st Century Skills*. Partnership for 21st Century Skills.

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 International Publishing.

EXPLORING THE TEMPORAL & SEQUENTIAL CHANGES OF SELF, CO AND SOCIALLY SHARED REGULATION IN ONLINE COLLABORATIVE LEARNING ENVIRONMENT: A LEARNING ANALYTICS APPROACH

Muhammad Azani Hasibuan
The University Western Australia
muhammad.hasibuan@research.uwa.edu.au

ABSTRACT: Over the past decade, many studies have attempted to explore the role of students' regulatory abilities in improving the quality of students' learning experience, especially in a collaborative environment. One of the issues in these research areas is how to measure and understand the temporal and sequential changes of the student regulations in every stage of collaboration. Temporal changes refer to the transient changes of student regulatory level in a period of regulation. The sequential changes refer to order of the type of regulation processes that happened in a period of collaboration. Answering this question could inform educational technology researchers and educators to support collaborative learning using information technology. Although there are already several studies that explore changes in student regulation, those studies were time-consuming because in manual observation and coding of student regulation. In the context of Computer-Supported Collaborative Learning (CSCL), this process could be done automatically by employing learning analytics. The present study attempts to extend existing research boundaries by exploring how students' self, co, and socially-shared regulation interact and influence each other, and how students' regulation changes in a CSCL course in software engineering. Additionally, this study aims to develop a new approach to understand the changes and development of students' regulations by employing learning analytics to capture student regulations

Keywords: Learning analytics, self-regulated learning, Co and shared self-regulated learning, computer-supported collaborative learning

1 INTRODUCTION

In the era of the digital economy, the ability to work collaboratively in a global working environment is a mandatory skill for every software engineer (Trilling & Fadel, 2009; Whitehead et al., 2010). In response to the need for this skill, Computer Science (CS) educators have already developed several initiatives to promote collaborative learning in several courses. One of the efforts is utilizing Information Communication and Technology (ICT) to support collaborative learning, also known as Computer-supported Collaborative Learning (CSCL) (Kirschner et al., 2013). The examples of these technologies, among others Learning Management System (LMS), Wiki, Social Network. In addition to these technologies, computer science educators also utilize specific collaboration tools for software engineering projects like *GitHub* or *Bit bucket*.

In collaborative learning environments, including CSCL, the achievement of learning objectives depends on the level of the students' regulations also known as self-regulated learning (SRL) (Järvelä, Järvenoja, et al., 2019). SRL is described as a metacognitive ability of the learner to control their emotion, cognition, and behaviour to acquire knowledge or skills (Zimmerman, 1989). SRL can be viewed as an aptitude behaviour of students or as a series of events that occur in a real learning environment. Capturing and measuring a student's SRL as an aptitude is relatively easy because the researcher only collects the perceptions of the students toward their SRL behaviour. In the context of

SRL as a series of events, the researchers must take significant effort to observe student activities in a real learning environment. Measuring SRL as events will depend heavily on the context of the learning environment because the different context will have distinct learning activities.

In the context of CSCL, most of the studies measure the students' regulations by incorporating these two views. These typical studies provide triangulation of SRL measurement that helps to get a holistic understanding of how regulations work in collaborative learning. Previous studies of student regulations in CSCL have shown that student regulations are changing over the period of collaboration (Järvelä, Järvenoja, et al., 2019; Molenaar & Chiu, 2015). Most of these studies happened in the context of the blended collaborative learning environment where the majority of the interaction of the students was in the face to face mode. When capturing student activities, most of these studies used visual observation through video recording. Data collected by this method are difficult and tedious to analyze because the researchers have to identify and code each of SRL activities manually.

What remains unexplored is how the students' regulations (self, co and shared) change and develop in the context of complex online collaboration as in a software engineering course, where the majority of the student's activity happens in an online environment. In this context, the process of capturing SRL events is relatively easy because all the student's actions are recorded automatically by the online learning or collaboration platform in the form of web usage logs. This type of data requires a new method of analysis, and the previous studies suggest using learning analytics –application data mining and machine learning for educational data-as an approach to analysis. Hence, the current research pushes existing research boundaries by exploring how students' self, co, and socially-shared regulation interact and influence each other and how students' regulation changes in a CSCL course in software engineering. Besides that, this study also tries to develop a new approach by employing learning analytics to capture student regulations

2 MEASURING SELF-, CO-, AND SOCIALLY-SHARED IN CSCL AS TEMPORAL AND SEQUENTIAL PHENOMENA

Measuring the self-regulatory process has become an interest for many researchers (Järvelä & Hadwin, 2013). Several instruments have already been developed to capture student regulation. Philip et al. (2000) identified two groups of instruments based on the perspective of regulation. There are two types of regulation perspective, self-regulation as aptitude and self-regulation as an event. Within the aptitude perspective, there are several ways to measure the self-regulation, like self-report questionnaire, structured-interview, and teacher judgment. From all of these instruments, the self-report questionnaire is the standard tool used to measure self-regulation. While in the event perspective, the measurement of self-regulation is done through a think-aloud protocol, temporal analysis, microanalysis, observation, and teacher judgement (Molenaar & Chiu, 2015; Sobocinski et al., 2017; Zimmerman, 2008).

Although there is a significant development in self-regulation measurement research, there is limited research on measuring self-regulation as a shared metacognitive process (Co- and Socially-shared regulation), particularly in the context of CSCL. Also, there is more limited research measuring the changes of the regulatory process throughout every stage of collaboration (Dindar et al., 2019; Järvelä et al., 2020; Järvelä & Malmberg, 2015; Malmberg et al., 2017). Researchers have tried to capture the temporal and sequences of Self, Co- and socially-shared regulation using various approaches. For example, a study by Malmberg et al.,(2017) used the videotape to record the student activity in a collaborative activity, and then code the activity manually. Similar to Malmberg er al., Sobocinski et al., (2017) also use videotape as the data observation. However, they combine it with the insight that gathers the digital traces of student activity using process mining. Another study from Dindar (2019) tries to capture the temporal changes of SRL by combining data from electro dermal activity with self-report. The results from these studies show that the student-regulations are changing in different stages of collaboration.

What remains unclear from the previous studies is how student-regulation changes and develops in the context of a complicated CSCL course, such as a Software-Engineering Project. Self-regulation and Co and Socially-shared regulation, are domain-specific (Greene et al., 2015). In other words, the context has a crucial influence on the students' regulations and how it will be measured.

3 LEARNING ANALYTICS FOR MEASURING THE CHANGES OF SELF-REGULATED LEARNING IN CSCL

The digital traces yielded in an online learning or CSCL environment are a valuable treasure for the researcher, especially in the area of self-regulation measurement. This abundance of data requires a new approach to analysis; one emerging approach is data mining. The research area that focuses on utilizing data mining in the educational context is known as learning analytics (LA) or Educational Data Mining (EDM) (Phil long, 2016).

In the context of measuring the changes of SRL or SSRL in a CSCL, learning analytics has the potential to increase the efficiency of research since it can capture and classify the student activity in CSCS automatically. Compare the se of learning analytics to the study conducted by Malmberg et al. (2017) who used videotape to observe the student activity in order to capture the regulatory process of the student; the LA solution will help to reduce the work load of the researcher.

Regarding its validity, Li et al. (2020) provide a methodological foundation to use digital traces data as a source to self-regulated learning. Their study shows that there is a correlation between student digital-traces data and performance. The digital traces data represent the time-management and effort-regulation construct of SRL. Learning analytics also has a strong dependency on data quality (Farrell, 2018). There are many data available in an online learning environment, and most of them are not relevant for measuring SSRL. For this reason, we need a measurement model that will translate SRL variables into evidence that can be found in CSCL. Unfortunately, there is no such model available that explicitly maps the Self, Co, and Socially-shared variables to digital traces in an online collaborative environment for a software engineering project.

4 THE RESEARCH GAP

As discussed earlier, there is a limited understanding of how Self-, Co- and Socially-shared regulation change and develop in collaborative learning, especially in the domain of Software Engineering where most of the interaction happens remotely through a collaborative learning platform. The understanding of how students' regulations change and develop is necessary as a foundation to design the tools or intervention to support collaboration in online learning. Previous studies only examined the changes of individual regulation and social regulation where the student interaction happen in an offline environment (Dindar et al., 2019; Järvelä, Malmberg, et al., 2019; Molenaar & Chiu, 2015; Sobocinski et al., 2017). Understanding the Self-, Co- and socially-shared regulation in the context of an online environment requires a new method for both data collection and analysis. One promising approach that suggested by several studies is Learning analytics. While several studies have utilized learning analytics to predict student performance in online learning, few studies that utilized it to analyses the student' Self-, Co- and socially-shared regulation, especially in the domain Software Engineering.

5 THE RESEARCH QUESTIONS

The current research will explore how the student' Self-, Co- and socially-shared regulation change and develop based on temporal data trajectory from online collaboration tools. This research will use mixed exploratory sequential research design. The following are the main aims and related questions of this research:

1) To develop the feature mapping framework for measuring students' regulation based on digital traces from an online collaboration platform. The following are related research questions for this objective:
    - Research Question 1 (RQ1): In what extent the existing learning analytics framework can identify the theoretical construct of students' regulations (self, co and social) based on the evidence that available in an online collaborative learning environment? If there is a limitation, what kind of framework that can map that can fulfil the gap?
2) To develop a theory-driven learning analytics learner model that can represent the level of students' regulations (Self, Co and Shared) based on their digital traces from an online collaborative learning environment. To achieve this objective, the following question will be addressed:
    - Research Question 2 (RQ2): What kind of learning analytics model can classify and assess accurately the level of students' regulations based on their digital traces? How accurately does the model classify and assess students' regulations?
3) To provide new theoretical understanding based on multimodal learning analytics approach regarding how students' Self, Co, and Socially-shared regulation changes and develops in complex computer-supported collaborative learning. The related question for this aim is:
    - Research Question 3 (RQ3): How do the students' regulations (Self-, Co, and Socially-shared) change and develop throughout every stage of collaboration in a complex online collaborative learning environment?

6 RESEARCH METHODOLOGY

The methods of the study: To develop the learning analytics solution, several steps be followed:

1. Data acquisition: the first step of this study is to collect the data set. The sources of data sets come from students' activities that recorded in the following online collaboration learning platforms:

    - GitHub: This study will collect the log of student activity both individually and as a group.

    - Slack: From this platform, the log of student and the content of communication will be collected as data set.

- Moodle: This study will collect the log of the student activities as the data set.

2. Data pre-processing: this process will integrate and clean the data that come from different sources

3. Features selection: this study will use the feature-mapping framework as guidance to select the list of attributes from a complex online collaborative learning platform.

4. Develop the training data set: because this study will use a supervised algorithm, the availability of a training data set is mandatory. To develop the training data set, this study will recruit two experts as independent coders. The data set will be coded based on the type of regulation construct and the type of regulation. The validity of this data set will be measured by inter-rater agreement.

5. Model selection and evaluation: the next process is to select a suitable data mining algorithm associated with the type of data and the type of behavior that is to be observed. Three frequently major algorithms that used are process mining algorithm to capture the sequence of activity, the text-mining algorithm to extract information from student discourses, and the classification algorithm to classify the type of student regulation. This step also will evaluate the accuracy of the model using the specific evaluation method that associated to the type of algorithm

6. Visualization: the last step of this study is to develop visualization to communicate the analysis of students' level of self, co and shared regulation.

7 CURRENT STATUS OF WORK & RESULT

This study has just passed the proposal phase. At this time, this research is preparing a literature study on the existing learning analytics framework can identify the theoretical construct of students' regulations (self, co and social)

REFERENCES

Dindar, M., Malmberg, J., Järvelä, S., Haataja, E., & Kirschner, P. A. (2019). *Matching self-reports with electrodermal activity data : Investigating temporal changes in self-regulated learning*.

Farrell, T. M. (2018). Affordances of Learning Analytics for Mediating Learning. *PQDT - Global*, 321. https://search.proquest.com/docview/2154878433?accountid=17242

Greene, J. A., Bolick, C. M., Jackson, W. P., Caprino, A. M., Oswald, C., & McVea, M. (2015). Domain-specificity of self-regulated learning processing in science and history. *Contemporary Educational Psychology*, *42*, 111–128. https://doi.org/10.1016/j.cedpsych.2015.06.001

Isohätälä, J., Järvenoja, H., & Järvelä, S. (2017). Socially shared regulation of learning and participation in social interaction in collaborative learning. *International Journal of Educational Research*, *81*, 11–24. https://doi.org/10.1016/j.ijer.2016.10.006

Järvelä, S., Gaševi, D., Seppänen, T., Pechenizkiy, M., & Kirschner, P. A. (2020). Bridging learning sciences, machine learning and affective computing for understanding cognition and affect in collaborative learning. *British Journal of Educational Technology*, *0*(0), 1–16. https://doi.org/10.1111/bjet.12917

Järvelä, S., & Hadwin, A. F. (2013). New Frontiers: Regulating Learning in CSCL. *Educational Psychologist*, *48*(1), 25–39. https://doi.org/10.1080/00461520.2012.748006

Järvelä, S., Järvenoja, 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*(4), 425–441. https://doi.org/10.1007/s11412-019-09313-2

Järvelä, S., & Malmberg, J. (2015). *How Individual Self-Regulation Affects Group Regulation and Performance : A Shared Regulation Intervention*. https://doi.org/10.1177/1046496415591219

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 2018*, 101203. https://doi.org/10.1016/j.learninstruc.2019.04.004

Kirschner, P. A., Erkens, G., & Kirschner, P. A. (2013). Toward a Framework for CSCL Research Toward a Framework for CSCL Research. *Educational Psychologist ISSN:*, *48*(1). https://doi.org/10.1080/00461520.2012.750227

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 and Education*, *104*, 18–33. https://doi.org/10.1016/j.compedu.2016.10.001

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

Malmberg, J., Järvelä, S., Järvenoja, H., & Panadero, E. (2015). Promoting socially shared regulation of learning in CSCL : Progress of socially shared regulation among high- and low-performing groups. *Computers in Human Behavior*, *52*, 562–572. https://doi.org/10.1016/j.chb.2015.03.082

Molenaar, I., & Chiu, M. M. (2015). Effects of sequences of socially regulated learning on group performance. *ACM International Conference Proceeding Series*, *16-20-Marc*, 236–240. https://doi.org/10.1145/2723576.2723586

Panadero, E. (2017). A Review of Self-regulated Learning : Six Models and Four Directions for Research. *Psychological Bulletin*, *8*(April), 1–28. https://doi.org/10.3389/fpsyg.2017.00422

Phil long, georger siemens. (2016). Penetrating the Fog. *Nursing Science Quarterly*, *29*(3), 237–240. https://doi.org/10.1177/0894318416647779

Sobocinski, M., Malmberg, J., & Järvelä, S. (2017). *Exploring temporal sequences of regulatory phases and associated interactions in low- and high-challenge collaborative learning sessions*. 275–294. https://doi.org/10.1007/s11409-016-9167-5

Trilling, B., & Fadel, C. (2009). 21st century skills: Learning for life in our times. In *21st century skills: Learning for life in our times.* Jossey-Bass.

Whitehead, J., Mistrík, I., Grundy, J., & van der Hoek, A. (2010). Collaborative Software Engineering: Concepts and Techniques. In I. Mistrík, J. Grundy, A. Hoek, & J. Whitehead (Eds.), *Collaborative Software Engineering* (pp. 1–30). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-10294-3_1

Zimmerman, B. J. (1989). A Social Cognitive View of Self-Regulated Academic Learning. *Journal of Educational Psychology*, *81*(3), 329–339. https://doi.org/10.1037/0022-0663.81.3.329

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, *45*(1), 166–183. https://doi.org/10.3102/0002831207312909

Classification and Clustering in Innovation-Based Learning: from Pilot Study to Practice

Lauren Singelmann
North Dakota State University
Lauren.N.Singelmann@ndsu.edu

ABSTRACT: Advancing our society will require both innovation and innovators, but traditional educational models do not focus on building innovative thinkers. A new model for instruction, Innovation-Based Learning allows students to innovate real-world solutions to some of society's most vexing problems while also learning course concepts. In this model, students use a learning management system platform (MOOCIBL) to track their learning and project progress. This extensive log data provides a comprehensive look into how students approach the complex problems of innovation. Classification and clustering models have been used to predict and better understand how undergraduate and graduate engineering students approach the process of innovation in the course. Past research during a pilot study demonstrated the efficacy of classification and clustering in predicting student success. Current work is being done to develop and implement a framework that allows for comparisons between students across different cohorts. Ultimately, this research may expand learning analytics in real-world problem-solving and better support instructors to help their students develop innovation skills to make a real-world impact.

Keywords: Complex problem solving, innovation, classification, clustering

1 INTRODUCTION

To make advancements in areas ranging from healthcare to the environment to infrastructure, we need to help students learn to excel in complex problem solving (National Academy 2004). To better prepare undergraduate and graduate engineering students in becoming critical and creative thinkers, our group developed a new education model: Innovation-Based Learning (IBL). Rather than being assessed on homework, tests, and quizzes, student learning is assessed on their ability to deliver impactful innovations for real-world problems. Students keep track of their project progress (and, most importantly, what they are learning along the way) in an online portal called MOOCIBL. Many students have thrived in this model; outcomes have included students publishing papers, submitting invention disclosures, and even creating companies (Singelmann 2020a), but it can be challenging for an instructor to scale up this model and still provide student support. Therefore, this work aims to use learning analytics and educational data mining to better understand how students approach complex problem solving, with the ultimate goal of better supporting both teachers and instructors in this model.

1.1 Goals of the Research

This work consists of two main phases: 1) an exploratory pilot study, and 2) development of a framework to improve the analysis process. Each of these phases has a variety of research questions which are summarized in Figure 1. The exploratory pilot study consisted of two components: A) using interpretable classification models to determine if the data collected can predict student success in creating an innovation, and B) using a clustering model to see how students are grouped and what characteristics define each of those groups. These models converted student text to TF-IDF matrices, meaning the models were trained by analyzing the frequency of words used. After the completion of the pilot study, these classification and clustering algorithms were applied to a new cohort, but the algorithms were not found to be sufficiently generalizable for the new cohort. Therefore, the goal of Phase 2 was to develop a framework that categorizes student text; these categorizations allow students from both cohorts to be compared (even if their class vocabularies are not exactly the same). Phase 2 consisted of three components: A) developing a framework that accurately, generally, and simply represents student data, B) developing a text classifier to automatically group student text into the categories of the framework, and C) using the new categorized data to learn more about how students approach the complex problem-solving process.

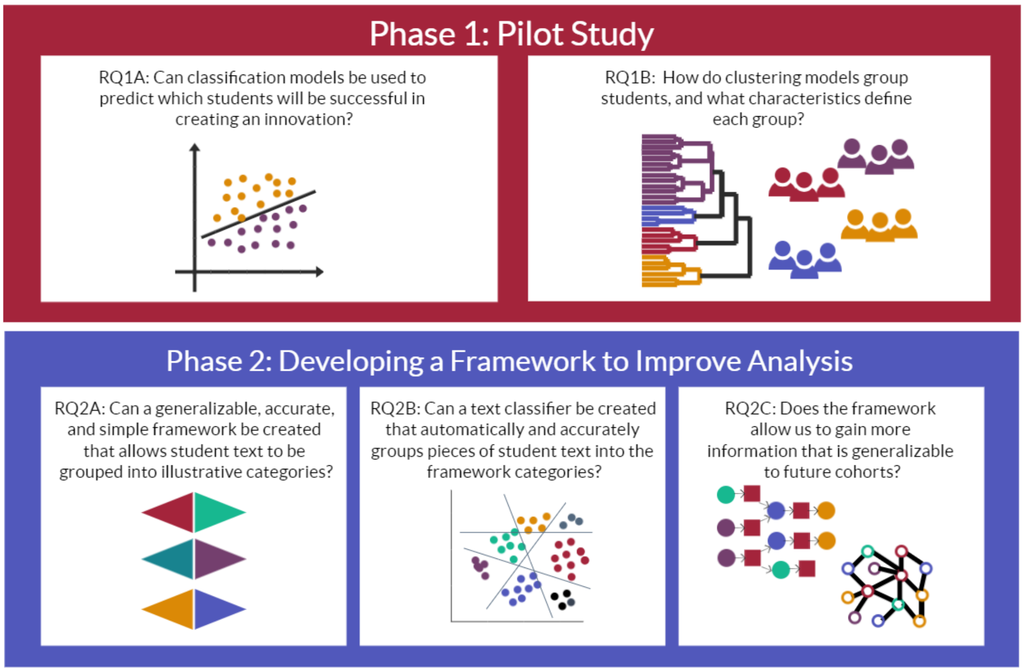


Figure 1: The 2 Phases include the pilot study and creating a framework to improve analysis. Each phase has research sub-questions.

2 BACKGROUND

An emerging area of learning analytics is exploring student problem solving, but measuring complex problem solving such as innovation requires a complex strategy (Buckingham Shum 2016). An approach that is based in learning sciences literature is needed, especially when working with complex ideas and relationships. For example, Zhang and Chen (2016) measure and explore student epistemic agency by analyzing how students highlight, connect, and annotate ideas in an online platform. Giabbanelli et al. (2019) compare student mind maps with expert mind maps to assess the ability to work on ill-structured problems. Martin et al. (2016) use trace log data from an interactive platform for learning about media literacy and sharing work. Other strategies range from text mining to object/body tracking, to looking at student programming (Blikstein 2016).

3 PROPOSED SOLUTION

MOOCIBL tracks complex problem-solving using learning tokens, with each token being a topic that a student needs to learn to advance their project. Students have the freedom to learn content and skills are applicable, but they also have a framework for classifying tokens to show sufficient progress. When a token is added, edited, or deleted, these actions are logged and can be analyzed using various machine learning techniques. As defined in the methods, previous work has included using supervised learning (classification) and unsupervised learning (clustering). While some other methods for measuring the ability to work on complex problems use specific prompts and situations, MOOCIBL is unique in that it allows students to work on significant problems that have not been solved. Nevertheless, this freedom is balanced with consistency established through tracking the learning process with tokens. Ultimately, the use of classification and clustering on student-developed text in IBL stands out in two ways. First, it supports an educational model where student learning is assessed on ability to innovate real-world solutions. Secondly, it aims to not only help us better understand the innovation process, but also predict student success, which is uncommon in other complex problem-solving approaches (Buckingham Shum 2016).

4 METHODS

4.1 Phase 1

4.1.1 Determining the Efficacy of using Classification on Innovation-Based Learning Data

By exploring two different types of feature sets (quantitative and text) and three different algorithms (support vector machine K-nearest neighbor, and logistic regression), the classification model was optimized for the pilot dataset. The final model uses the text that students wrote about their learning and creates a support vector machine. By using a linear kernel, this model is also interpretable, meaning the words that differentiate between top performers and lower performers were able to be extracted. All models were evaluated using ten-fold cross validation, which gives a measure of how well the model might perform on new datasets. The final model had accuracy of about 90% and a ROC AUC score of 0.95 (Singelmann 2020c).

4.1.2 Determining the Efficacy of Using Clustering on Innovation-Based Learning Data

Clusters were discovered using agglomerative clusters, and these clusters were then named by observing what words came up most in each cluster. The clusters were named *Innovators, Learners, Surveyors,* and *Surface Level,* and definitions for each cluster were developed. Instructors of the course then predicted what cluster each student would fall into, and these results were compared with the model. Kohen's Weighted Kappa was 0.608, meaning there was moderate agreement. These clusters were then mapped to the level of Bloom's Taxonomy and Webb's Depth of Knowledge that each cluster was able to meet (Singelmann 2020b).

4.2 Phase 2

4.2.1 Creation of the Framework

In order to create an appropriate model for the data, the alternate templates strategy was used, which is common in qualitative research dealing with complex process data (Langley 1999). Alternate templates strategy consists of qualitatively reviewing the data and the literature, developing an appropriate model, assessing how well that model fits your data, and continuing to reiterate that process until a final framework is selected. An ideal framework should balance three things: simplicity, generality, and accuracy (Langley 1999). The developed framework is general enough to fit any student from any cohort, it is accurate enough that the interrater reliability (Cohen's Kappa) is greater than 0.6, and it is simple enough that it contains only seven framework components.

4.2.2 Development of a Text Classifier

After the final framework is completed, over 750 pieces of student text were classified into the framework components by the author. The author used context from the rest of the log, information in the literature, and understanding of the student group and project to inform the categorizations. Since this process is very time-consuming, component two of Phase 2 was to assess the viability of a classifier model that could complete the task in a matter of seconds for future student data. Eight different models were trained using a combination of four different algorithms (logistic regression (LR), k-nearest neighbors (KNN), random forest (RF) and support vector machine (SVM)) and two different feature types (unigrams, and combined unigrams and bigrams). Five-fold cross-validation was used when measuring performance, and accuracy, Cohen's Kappa, and the F1 score were calculated for each of the eight models. Accuracy looks simply at the percentage of tokens where the class predicted by the model matched the class chosen by the human rater. Because there is some subjectivity in categorizing the objectives, Cohen's Kappa was also used as a performance metric; Cohen's Kappa is commonly used to score interrater reliability in qualitative research. Rather than assuming that the human classification is always correct and that the model is trying to match the human classification perfectly, Cohen's Kappa assumes that the goal is to have agreement between the human and the model. F1 score is the average between the precision and the recall of the model.

4.2.3 Analysis with Framework

The third proposed component is to use the categorized data to analyze student behavior. Rather than treating all text features equally, students can be compared at the category level. For example, what differentiates successful students from struggling students in the "Survey" stage? In addition, by categorizing the data, analysis methods such as process analysis and network analysis can be used to explore how individual students and teams approach the innovation and problem-solving process. This component is intentionally left open-ended; research questions can adjust based on findings and questions that the researchers and instructors of the course might develop.

4.3 Ethical Considerations

MOOCIBL's main goal is to provide support for students and instructors working within Innovation-Based Learning or other open-ended problem-solving environments. However, care is to be taken to ensure that all students are supported. Because predictive models are a simplification of the total system, sometimes certain groups of students can be consistently misclassified (Corrin 2019). Using interpretable models, incorporating human expertise, and validation of results using findings from the literature are three of the strategies that MOOCIBL uses to work towards equitable solutions for all learners. Some machine learning methods are "black-box models", meaning it is unclear what information the model is using to make its decisions (Romero 2013). MOOCIBL uses only models that allow for knowledge discovery, meaning the important features are extracted to understand how the model is making its conclusions. Instructors and researchers pay careful attention to the algorithms' outputs and work to validate them with cognitive science and engineering education literature findings.

5 CURRENT STATUS

The data from the pilot semester has been analyzed and results have been published as detailed in Section 5.1. The work from these studies has shown that there is sufficient efficacy of both methods to be able to continue with future studies. However, the exploratory models performed poorly on the new cohort. Therefore, a framework that allows for comparisons between students in both cohorts (and future cohorts) was created. A summary of this work is detailed in Section 5.2.

5.1 Papers Published

Four conference papers have been published about the past work. *"Design and Development of a Machine Learning Tool for an Innovation-Based Learning MOOC"* detailed the methods for how data would be collected within the platform (Singelmann 2019). *"Student-Developed Learning Objectives: A Form of Assessment to Enable Professional Growth"* discussed the pedagogical decisions behind Innovation-Based Learning (Singelmann 2020a). *"Predicting and Understanding Success in an Innovation-Based Learning Course"* explored the performance of the classification models and detailed how feature extraction could help lead to better understanding of the innovation process (Singelmann 2020c). *"Innovators, Learners, and Surveyors: Clustering Students in an Innovation-Based Learning Course"* detailed how the clusters were discovered and aligned each of the clusters to various learning taxonomies (Singelmann 2020b).

5.2 Experiments in Progress

Currently, the second cohort of students has completed the Cardiovascular Engineering course. Because the exploratory models trained with the 2019 data did not have strong performance on the 2020 data, a framework has been created to help group student text into illustrative categories: survey, define, explore, solve, develop, share, and environment. This framework focuses on the convergent and divergent behaviors that students use in complex problem-solving and innovation (Wolf 2009, Van de Ven 2017); survey, explore, and develop are divergent behaviors where the students are exploring the problem and solution space, and define, solve, and share are convergent behaviors where the students are making decisions and refining their problem statement and solution.

A text classifier has also been created to automatically group student text into these categories; the text classifier and the human classifier have substantial agreement (Cohen's Kappa > 0.7). Current work includes exploring which framework components are most important in predicting student success and understanding student strategies in the course, but support is needed to determine what methods are the most appropriate and have the most promise in such a unique context. The ultimate goal is to use the findings to create specific student-centered interventions that can be applied and measured for students taking the course in 2021. By improving student support in this course, we are equipping students to solve real problems with real value, improving both engineering education and the world around us.

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.

Buckingham Shum, S., & Crick, R. (2016). Learning Analytics for 21st Century Competencies. *Journal of Learning Analytics*, 3(2), 6–21.

Corrin, L., Kennedy, G., French, S., Buckingham Shum, S., Kitto, K., Pardo, A., West, D., Mirriahi, N., & Colvin, C. (2019). The ethics of learning analytics in Australian higher education. *Accessed online,* 26.

Giabbanelli, P., Tawfik, A., & Gupta, V. (2019). Learning Analytics to Support Teachers' Assessment of Problem Solving: A Novel Application for Machine Learning and Graph Algorithms. *Accessed online.*

Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management review*, *24*(4), 691-710.

Martin, C., Nacu, D., & Pinkard, N. (2016). Revealing Opportunities for 21st Century Learning: An Approach to Interpreting User Trace Log Data. In *Journal of Learning Analytics,* 3(2), 37-87.

National Academy of Engineering, U. S. (2004). *The engineer of 2020: Visions of engineering in the new century.* Washington, DC: National Academies Press.

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* 3(1), 12–27.

Singelmann, L., Alvarez Vazquez, E., Swartz, E., Pearson, M., & Striker, R. (2020). Student-Developed Learning Objectives: A Form of Assessment to Enable Professional Growth. In *2020 ASEE Annual Conference.*

Singelmann, L., Alvarez Vazquez, E., Swartz, E., Pearson, M., Striker, R., & Ewert, D. (2020). Innovators, Learners, and Surveyors: Clustering Students in an Innovation-Based Learning Course. In *2020 Frontiers in Education Conference.*

Singelmann, L., Alvarez Vazquez, E., Swartz, E., Striker, R., Pearson, M., & Ewert, D. (2020). Predicting and understanding success in an innovation-based learning course. In *Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020)* (pp. 662–666).

Singelmann, L., Swartz, E., Pearson, M., Striker, R., & Alvarez Vazquez, E. (2019). Design and Development of a Machine Learning Tool for an Innovation-Based Learning MOOC. In *2019 IEEE Learning With MOOCS (LWMOOCS)* (pp. 105–109).

Van de Ven, A. H. (2017). The innovation journey: you can't control it, but you can learn to maneuver it. *Innovation*, *19*(1), 39-42.

Wolf, K. M., & Mieg, H. A. (2010). Cognitive determinants of the success of inventors: Complex problem solving and deliberate use of divergent and convergent thinking. *European Journal of Cognitive Psychology*, 22(3), 443-462.

Zhang, J., & Chen, B. (2016). Analytics for Knowledge Creation: Towards Epistemic Agency and Design-Mode Thinking. In *Educational Theory and Practice Faculty Scholarship,* 21.

Predicting Student Performance in an Accounting Course with Usage Data from Gamified Learning App

Julian Langenhagen
Accounting Department, Goethe University Frankfurt, Germany
langenhagen@econ.uni-frankfurt.de

ABSTRACT: We developed a gamified mobile learning app to support the students' learning process in an undergraduate management accounting course at a large public university in Germany. The app allows students to test their learning achievements by answering numerous quiz questions. The goal of our research project is to predict the students' performance in the final exam with the usage data from the app. Furthermore, we investigate whether the prediction model can serve as an early predictor during the semester. Additionally, we examine the portability of the prediction model to future semesters, especially when the teaching method switched from face-to-face to fully online due to COVID-19. Although the preliminary results indicate that the usage data has the potential to predict student performance in the final exam, the predictive power needs to be increased. Therefore, further machine learning algorithms and innovative performance measures for future work are discussed.

Keywords: Learning Analytics, Gamification, Mobile Learning, Accounting, Portability

1 INTRODUCTION

The setting of our research project is as follows: We developed a mobile learning app for an undergraduate accounting course at a large public university in Germany. The course is compulsory and ought to be taken in the third semester of the bachelor's program. The course is taken by approximately 600 students per semester and consists of a weekly lecture, a biweekly exercise, and biweekly tutorials (5 separate meetings in small groups). The content of this course includes the basics of cost accounting as well as a summary of their significance and classification in the management accounting context. The primary learning material consists of a slide deck, a collection of exercises (with solutions), and a trial exam (all available as PDF files). In the evaluations of earlier semesters, students often complained that there were no contemporary possibilities to learn the subject matter. Therefore, we decided to develop an additional learning tool in the form of a smartphone app named BaccUp[1], which was launched in the summer semester of 2019. The use of BaccUp is voluntary and no extra credits or advantages for the final exam can be earned by collecting points in the app. The tool is available both via a web version and as an app in the Google Play Store and the Apple App Store. The core element of the app is a database with over 550 questions that covers all nine chapters of the course. In addition to the question types single and multiple-choice, there are also sorting and cloze text tasks. The app can be used in three different modes: The chapter mode can be used to answer specific questions about a single chapter. As soon as a student has mastered the problems of one

[1] The name of the tool consists of the abbreviation of the course (BACC) and the word "up".

chapter, the next chapter is unlocked. In random mode, questions are randomly selected from the chapters that have already been unlocked in chapter mode. In the third mode, the so-called Weekly Challenge, users can compare themselves with other students. Once a week, they have the opportunity to answer 25 questions randomly selected from the chapters already covered in the lecture. The results are subsequently displayed in a weekly and a semester ranking. For good performances in the Weekly Challenge and other learning achievements, students can earn so-called badges, which are then displayed in their account under their self-chosen username (see Figure 1). By answering questions (regardless of the mode), students also earn learning points and thus increase their learning level. The progress display of the individual chapters shows students how well they currently master a particular topic (see Figure 2). The app has been specifically designed to complement the existing course and is not intended to replace other learning materials such as the slides or the collection of exercises. The app contains an individual explanation for each question which is displayed if a wrong answer is given. Thus, students can work their way through the catalog of questions independently of time and place and eliminate any gaps in their understanding without having to rely on the presence of the lecturers. This is an essential value-added for the students, especially in such a large course with approximately 600 students per semester.

Table 1: Number of students across semesters.

	SS18	WS18	SS19	WS19	SS20	WS20
Exam	590	671	**575**	**648**	**616**	?
App	NA	NA	**561**	**595**	**447**	?
Survey	153	250	127	156	114	?
Teaching	face-to-face		face-to-face		online	

Focus of this study is marked in bold

The collected data for our research project consists of three different sources: (1) usage data of the app users since the first semester it was used (summer semester 2019), (2) survey data, and (3) exam scores (see Table 1 for the corresponding sample sizes). Since the summer semester 2018, a survey accompanying the lecture has been conducted to capture and compare certain characteristics of the two student cohorts that were not able to use the app. The survey was divided into four sections: “Time and type of use” (e.g., to which extent the different learning materials were used or where the students typically learn), “Goals” (e.g., which goals were important regarding the course apart from passing the exam), “Satisfaction” (e.g., how satisfied the students were with different parts of the course) and “Other information” (e.g., sex, age, and prior knowledge). In the summer semester 2019, the survey was supplemented with questions on BaccUp (e.g., which elements were particularly motivating). The app data consists of details about the usage behavior of each student (e.g., time of use, performance (history) regarding every question, and earned badges). Compared to studies dealing with Massive Open Online Courses (MOOCs), the sample size of this study is small. However, for a face-to-face lecture at a university, the number of students is respectable, especially due to the increasing number over time as the project is still ongoing (Van Goidsenhoven et al., 2020). Nevertheless, there is an important limitation in our setting. In principle, we are not able to link the three data sets together at the student level, which for example means that we cannot link a student’s exam grade to their answers from the survey or their usage data from the app without their help. Due to data protection regulations, we were obliged to ask the students for their matriculation number in the app and survey on a voluntary basis to connect the data sets. As expected, not every student followed this request. Therefore, the data basis for analyses that require two or all three data sources is significantly reduced and affected by selection bias. For the analysis in this paper, only the app and exam data are used (marked in bold in Table 1). In this case, we have the following sample sizes: 229 students for the summer semester 2019, 242 for the winter semester 2019, and 180 for the summer semester 2020. However, even with this limitation, our data set has some promising properties for further research. At this stage, we already have three semesters of app usage and the data set is growing as the research project is still ongoing. According to our knowledge, this is the first study to examine the impact of a gamified mobile learning app on student performance in a large university course over the duration of several semesters. This is especially promising as the situation regarding COVID-19 lead to an exogenous shock. While in the years 2018 and 2019 the course was held face-to-face, in the summer semester 2020 it was converted into a purely online lecture. Apart from the launch of the app and the switch to online lectures, there were no teaching design changes over the course of the semesters. The lecturer and the learning materials remained constant as well as the design and the grading of the final exam. This unique setting could provide valuable insights into the impact of COVID-19 on higher education.

2 RESEARCH QUESTIONS AND RELATED WORK

For this research project, two research streams are important: gamification in education (e.g., Huang et al., 2020; Sailer & Homner, 2020) and learning analytics, especially the prediction of learning outcomes (e.g., Conijn et al., 2016; Van Goidsenhoven et al., 2020). The focus of this study is on the latter. Our main goal is to predict the outcome of the final exam with the usage data from the app. In many comparable studies the outcome variable is binary, only indicating whether a student failed or passed the exam (e.g., Malekian et al., 2020; Van Goidsenhoven et al., 2020). While some studies go one step further and aim to predict the grade (Conijn et al., 2016), we instead chose the points in the final exam as the outcome variable. We believe that this measure is more precise because (at least in our case) the same number of points can result in a different grade in a different semester not because the single student's performance is better or worse but because the grades are assigned based on the performance of the whole cohort of one semester. The exam design in the underlying course is constant over all semesters. It consists of three separate exercises with 30 points each resulting in a total potential score of 90 points. Studies on predictive learning analytics regarding face-to-face lectures are scarce compared to MOOCs because the latter often provides a more comprehensive data set due to the limitation on online teaching and learning. A recent study contributed to this research gap by examining a blended learning environment (Van Goidsenhoven et al., 2020). We would like to add the context of a mobile learning app by examining our first research question:

> **RQ1:** To what extent can the student performance in the final exam be predicted with the usage data of the app?

The term "usage data of the app" has deliberately been chosen very broadly, as the exact design of the corresponding measures offers a wide range of possibilities. Nevertheless, the goal is to find the best model to predict student performance in the final exam. A further question could be whether this model is portable to comparable courses (e.g., a course that uses the same learning app with different questions). Recent studies examined this question with regards to the portability of a prediction model across different courses (Conijn et al., 2016; Gašević et al., 2016). The main outcome of these studies is that the generalization of such models needs to be handled with care since the results significantly vary across courses and therefore the portability is low. In our setting, we do not have different courses. Still, we can answer the question of whether the results of a corresponding model stay the same over several semesters, especially when the teaching design switches from a face-to-face to an online course. Therefore, our second research question is:

> **RQ2:** What is the portability of a general model for predicting student performance across semesters, especially regarding changes due to COVID-19?

Given the usage data of the app can be used to predict the students' exam performance, a follow-up question would be whether the app could even be used as an early predictor during the semester. Although research has shown that the prediction accuracy increases over time (Tempelaar et al., 2015), in some cases not the data of the whole semester is needed to reach satisfying accuracies (Van Goidsenhoven et al., 2020). Our third research question adds to that research area:

> **RQ3:** What is the impact of considering the most recent activities prior to the exam in the models compared to only incorporating the activities from the beginning of the semester?

The goal of the following analyses is to provide baseline results to give a first idea regarding the beforementioned research questions. Further directions and possible refinements are discussed in more detail at the end of this paper.

3 PRELIMINARY RESULTS

In this section, the results of two analytic approaches are presented. In the first approach, the student's performance is predicted with a linear regression model. The dependent variable is the student's score in the final exam of the corresponding semester. For each semester, we run two regression models. The independent variable of the first model is the total amount of answered questions (TA) per student per semester. The intuition behind this model is that the more a student uses the app, the more he should learn and understand and consequently the more points he should earn in the exam. The second regression model contains the highest chapter reached (HC) as the independent variable. The app includes all nine chapters of the corresponding course and they need to be unlocked sequentially. A chapter needs to be completed to a certain degree to unlock the next one. Even after fully completing a chapter the included questions can still be answered to repeat the corresponding topics. This means that a student could remain in chapter one for the whole semester and answer the same question set over and over again. This strategy would result in a high amount of answered questions, but the highest chapter reached would equal one. Presumably, this strategy would not be the best to prepare for the final exam. Therefore, the highest chapter reached could be a promising choice as an explanatory variable. With the aforementioned analysis, RQ1 and RQ2 can be examined. To investigate RQ3, we split the usage data per semester into four parts. The first part contains the usage data for the first quarter of the semester, the second part contains the data for the first half, and so forth. With this split, the development of the predictive power of the models over time and therefore the answer to RQ3 can be examined. The results (R^2) of these linear regressions are shown in Table 2.

Table 2: Results from linear regression on total points in final exam (R^2).

Included time period	SS19		WS19		SS20	
	TA	HC	TA	HC	TA	HC
25%	0.008	**0.018**	0.008	0.014	0.010	0.017
50%	**0.026**	**0.056**	**0.023**	**0.036**	0.013	0.021
75%	**0.049**	**0.081**	**0.056**	**0.083**	0.010	**0.034**
100%	**0.095**	**0.138**	**0.063**	**0.115**	**0.029**	**0.080**

Models with statistically significant coefficients (P values of ≤0.05) are marked in bold

The first results indicate that highest chapter reached seems to be a better choice as an independent variable compared to total questions answered. The predictive power is higher at every point in time in every semester. Furthermore, it can be seen that the predictive power of every regression increases significantly with more data over time in almost every scenario. The predictive power (R^2) of the models with highest chapter reached as the independent variable range from 8.0% in the summer semester 2020 to 13.8% in the summer semester 2019, indicating limited portability over time. As there are no prior studies that are really comparable it is challenging to put these results in context. Results from a study predicting final exam grades with data from the learning management system range from 8% to 37% (Conijn et al., 2016). Therefore, the present results indicate that the app usage data can indeed be used to explain part of the variance of the student's points in the final exam. In the second approach of this analysis, the same data is used to predict whether a student fails or passes the exam with a logistic regression model. The results (AUC) are reported in Table 3.

Table 3: Results from logistic regression on binary exam outcome (AUC).

Included time period	SS19		WS19		SS20	
	TA	HC	TA	HC	TA	HC
25%	0.576	**0.580**	0.550	0.548	0.587	0.584
50%	**0.634**	**0.638**	0.570	0.569	0.608	**0.607**
75%	**0.670**	**0.670**	**0.622**	**0.623**	0.618	**0.625**
100%	**0.714**	**0.716**	**0.660**	**0.679**	**0.673**	**0.701**

Models with statistically significant coefficients (P values of ≤0.05) are marked in bold

In contrast to the results from the linear regression models, the logistic regression models with highest chapter reached as the independent variable are not better in every comparison with the ones using total questions answered. Although the differences are not very high, this relation is still true for the models with 100% of the data. Regarding the portability of the models, we can notice that the differences are by far not as large as the ones of the linear regressions. While the AUC scores do increase with more data over time, the results with the total data still are only slightly "acceptable" based on Hosmer and Lemeshow's (2000) interpretation. Therefore, none of these models should be used to predict whether a student passes the exam or not. Further additions to the existing models or other analytical approaches should be examined to improve predictive power.

4 CONCLUSION AND NEXT STEPS

On the one hand, the preliminary results show that the usage data of the gamified learning app indeed has the potential to predict student performance in the final exam. On the other hand, we also observe that predictive power needs to be increased to provide reliable results. The data from the second fully online semester will provide further insights into the question of portability. At the current stage, the usage data cannot reliably serve as an early predictor, but there are several potential areas for improving the predictive model. First, a more complex prediction algorithm could be used, e.g., a random forest classifier (Van Goidsenhoven et al., 2020) or a neural network (Okubo et al., 2017). As the literature on the predictive value of learning app usage data (in contrast to LMS data) is scarce, the development of innovative and meaningful performance measures could also be a promising avenue. Novel approaches could include the starting point of the app usage or the number of days the app was used, indicating a more regular learning style. Moreover, in the current analysis solely aggregate measures were used. Focusing on the sequential pattern of app usage could be another promising avenue of future research (Malekian et al., 2020). To sum up, the project provides an auspicious setting that could be used to close numerous research gaps.

REFERENCES

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2016). Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies*, *10*(1), 17–29.

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68–84.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. Wiley.

Huang, R., Ritzhaupt, A. D., Sommer, M., Zhu, J., Stephen, A., Valle, N., Hampton, J., & Li, J. (2020). The impact of gamification in educational settings on student learning outcomes: A meta-analysis. *Educational Technology Research and Development*, *68*(4), 1875–1901.

Malekian, D., Bailey, J., & Kennedy, G. (2020). Prediction of students' assessment readiness in online learning environments: The sequence matters. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 382–391.

Okubo, F., Yamashita, T., Shimada, A., & Ogata, H. (2017). A neural network approach for students' performance prediction. *Proceedings of the Seventh International Conference on Learning Analytics & Knowledge*, 598–599.

Sailer, M., & Homner, L. (2020). The Gamification of Learning: A Meta-analysis. *Educational Psychology Review*, *32*(1), 77–112.

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*, 157–167.

Van Goidsenhoven, S., Bogdanova, D., Deeva, G., Broucke, S. vanden, De Weerdt, J., & Snoeck, M. (2020). Predicting student success in a blended learning environment. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 17–25.

5 APPENDIX

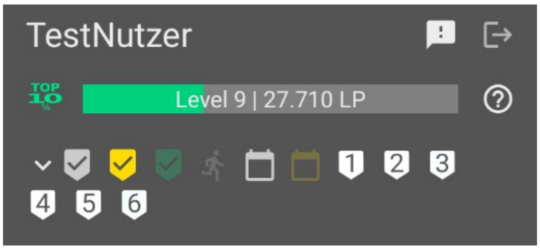


Figure 1: User view with points, current level and earned badges

Figure 2: Progress bars of different chapters

From Recipients to Learners: Unpacking Student Engagement with Learning Analytics in Higher Education

Yeonji Jung
New York University
yeonji.jung@nyu.edu

ABSTRACT: Students are an important target of learning analytics as its ultimate goal is to enhance learning, and most of the tools are based on student learning data. Despite growing attention to student-facing analytics in higher education, little is known about how students shape their learning practice with analytics beyond passive recipients of the tools and which area of support is needed to contextualize their analytics use, especially about sensemaking. My dissertation work addresses this issue by implementing, examining, and supporting student engagement with analytics in three phases: (a) problem analysis from literature review (completed), (b) design of different types of process-oriented analytics displaying student progress on online collaborative reading activities (in process), (c) implementation and examination of student use and sensemaking of analytics and their impact on course engagement (planned). The findings will contribute to conceptual basis of student analytics use and sensemaking and inform implications for supporting learning practices with analytics.

Keywords: Student-facing analytics, student analytics use, student analytic sensemaking

1 INTRODUCTION

Learning analytics has been shaping educational practices in higher education, supporting institutional decision making and instructional modifications to improve student learning experiences. While educators are frequently provided and examined with analytics use, students are a naturally important target of learning analytics. As most analytic tools are based on student learning data and developed with the ultimate aim of enhancing learning (Bodily & Verbert, 2017), the need for investigating how students benefit from and contextualize their own data is heightened (Prinsloo & Slade, 2016; Teasley, 2017). Despite its growing attention, student-facing analytics have been limitedly implemented, tending to focus on institutional-level adoption and intervention (Bodily & Verbert, 2007). Thus little is known about how students as end-users interact with and get informed by analytics for learning improvement in their context (Jivet et al., 2017). Going beyond the simple curiosity of what *analytics allow* students to do, a shift to focus on what *students do* with analytics in connection with their learning is needed (Prinsloo & Slade, 2016; Teasley, 2017). This approach could generate fresh insight into the field by positioning students as active agents engaging with analytics rather than passive recipients of the tools. Within practices of student analytics use, sensemaking activities are specifically important to examine as simple exposure to learning data does not always lead to actionable insights (Foster & Francis, 2019). Rather, this process requires complicated processing of student searching for relevant information, decoding its meaning, and reflecting on learning (Bodily & Verbert, 2007); otherwise, analytics may be just another tool seen as “a nuisance rather than an aid” (Klein et al., 2006, p. 71). Another confounding factor is that each student does not go through the same process of sensemaking. A single metric could be of different importance for individuals depending on contextual resources (e.g. course participation grade, peers’ progress, and instructors' emphasis on a certain material) and their learning characteristics (e.g. goals), which further leads to different changes in learning (Foster & Francis, 2019). However, little research has investigated the specific ways students engage with analytics and which area of support is needed to contextualize their analytics use considering factors impacting their use.

2 CONCEPTUAL FRAMEWORK & CURRENT KNOWLEDGE

2.1 Student use practices of learning analytics

While students have been engaged in receiving and responding to feedback through educational technologies, learning analytics tools are distinct in the kinds of information they display (e.g. both formative and summative information, reference frames for comparison) (Teasley, 2017) and the provision of interactive functions that students can take control over based on their needs (Bodily & Verbert, 2017). Such distinctiveness allows students to review their own data, keep track of learning progress, and make judgments on follow-up action, raising the need of investigating how analytics makes a difference in student learning (Jivet et al., 2017). This new area of focus is globally defined as *learning analytics use*. This term captures the human activity of working with these tools and the context of their use, which emphasizes the role of end-users in shaping this learning practice. The literature reported that despite some variations a high number of students accessed the analytics at least once when it was offered; however, once they accessed it, a small portion of them used the analytics frequently (Holman et al., 2020; Sansom et al., 2020). Four characteristics of analytics appeared to work as contributing factors shaping this use: (a) *context* (learning environments that implement student analytics use); (b) *information* (analytics and contents provided); (c) *analytic tools* (tools that deliver analytic information); (d) *support for use* (support through effective message design or provision of external supports for better analytics use and sensemaking).

Student visits to analytics were often made at a specific time related to the ***course context***; the most access to analytics was found just before the assignment deadline, exam periods, or timing to make decisions about grade (Holman et al., 2020). This suggests the role of learning context in prompting students to interact with analytics and processing analytics with contextually relevant resources (Klein et al., 2019). A specific type of ***information*** appeared to influence student perceptions of analytics use, showing mixed perceptions on outcome-oriented analytics and process-oriented analytics concerning its value to enhance learning (Lim et al., 2019; Wise et al., 2014), as well as reference frames provided that (de)motivated their analytics use (Bennett & Folly, 2019). Concerning ***analytic tools***, it was found that notification function embedded in personalized emails may help student visit analytics, but also the issue was raised about the optimized number of emails to be sent and the timing of its delivery in a way not to be perceived as annoying (Pardo et al., 2019). Despite a common assumption that students benefit from analytics use, for students to make use of analytics for learning, they need ***learning support*** to identify the value of analytics use and ways that they can meaningfully make sense of analytics to improve their learning (Klein et al., 2019). Most studies of student analytics use focused on effective design of message contents in different forms such as including prompting questions, motivational messages, and contextualized examples in the analytic tools (Pardo et al., 2019; Wise et al., 2014). In addition to different forms of learning support such as introduction workshops before use (Jivet et al., 2020) or interpretation guide (Chen et al., 2018; Jivet et al., 2020), a few cases took a step ahead to offer systematic support by framing analytics use as an integrated part of course activities that tied to course expectations and individuals' goals (Chen et al., 2018; Wise et al., 2014).

In addition to those four characteristics, student usage extent and timing of analytics were often found to be closely linked to ***student-related factors*** such as individuals' goals/motivation (Aguilar, 2018) and self-regulated skills (Kia et al., 2020).

2.2 Student sensemaking of learning analytics

An important distinction should be made that learning analytics use aims to go beyond adoption and consumption, which speak to behavioral questions about what students did or do with analytics. Rather, this concept should include learning to learn, inquiring into how analytics empower students to be aware of their progress and initiate planning follow-up strategies for improvement (e.g. "seeing the information of my unproductive study habits, I started thinking about how to efficiently manage my schedule" instead of "I used analytics") (Teasley, 2017). These activities are often framed by the notion of ***sensemaking*** (Lim et al., 2019). Sensemaking is a core part of student use of data; as data does not speak for itself, analytics use requires students to actively participate in deciphering the meaning of information and figuring out what to do to improve learning (Wise et al., 2014). In the field of learning analytics, sensemaking generally refers to a process of engaging with data in which users interpret the information provided and translate it into actionable insights (Lim et al., 2019). While the field has yet to draw specific dialogue around sensemaking, broader contexts of psychology and human-centered interaction (Klein et al., 2006; Pirolli & Card, 2005) attempted to build an understanding of sensemaking process, suggesting commonalities of four components: (a) information seeking for exploration, (b) information decoding for definition, (c) information evaluation for meaning making, and (d) changes in behaviors/thoughts in response to sensemaking.

The majority of studies probing student analytic sensemaking focused on ***information seeking***, identifying which information got attention the most. When given a variety of information in the tools, students commonly began by looking at the outcome-oriented information, along with peer values of the scores in comparison with their values (Kia et al., 2020). The second component of sensemaking, ***information decoding***, is critical for students to take steps forward sensemaking; however, only a few studies delved into how students attempted to decode the information, finding that students favored a particular metric that was easy to process and comprehend how it was calculated (Wise et al., 2014) and generated different interpretations of the same metrics depending on the reference frames (Aguilar, 2018). While it was generally assumed that data literacy would be a major challenge in accurately decoding analytics (Sansom et al., 2020) and thus provision of support for analytic interpretation is important, little work has been done to probe how and what kinds of supports helped students effectively decode the information. Given one opposite finding that most students were able to describe what the analytics indicated (Corrin & De Barba, 2015), it may be suggested to specify which areas of the decoding practices require additional interpretation support. The third component of sensemaking, ***information evaluation***, is a key part of student sensemaking of analytics, empowering students to connect the information decoded to their learning experiences. Several studies showed limited but promising evidence that students constructed the meaning of the analytics by facilitating reflection (Bennett & Folly, 2019; Corrin & De Barba, 2015), linking learning strategies with performance (Wise et al., 2014), making comparisons with peer-referenced frames (Bennett & Folley, 2019; Corrin & De Barba, 2015; Wise et al., 2014). Despite common use of retrospective strategies, only a few proactive attempts were identified by students often along with a lack of confidence in assessing which parts of learning could be improved and what ways students could make changes (Corrin & De Barba, 2015; Wise et al., 2014). This challenge may impede learning to learn with analytics, raising the need to support students in identifying the relevant meaning of analytics and situating it into their experiences. The last component of sensemaking, ***changes in behaviors/thoughts***, is a responsive part to the previous components and a key piece to make their use meaningful and make a difference in learning. Despite the common findings that providing analytics as an intervention had a positive effect on learning (Foster & Francis, 2020; Pardo et al., 2019), in-depth analysis found that changes made directly responding to analytic sensemaking did not always occur. Students may encounter challenges in deciding what to do next based on interpretation (Corrin & De Barba, 2015), even when they identified which part of their learning needs improvement (Wise et al., 2014). They also made inadvertent changes, resulting from their misunderstanding of the analytics (Klein et al., 2019; Wise et al., 2014). While this suggested that four components of sensemaking connect to and influence each other, little is known about the details of how a series of sensemaking activities may lead to a particular learning change.

3 RESEARCH GOALS & QUESTIONS

My dissertation work addresses this issue by going beyond student experiences as receivers of analytics to unpack detailed practices of how students engage with data-informed learning. The aim of my doctoral work is threefold: (a) to understand the current student practices of analytics use and sensemaking from the existing literature to guide the next phases; (b) to co-design different types of process-oriented analytics with students that can support behavioral, social, and conceptual engagement in online collaborative reading activities (general one in a report form vs. personalized one that offers a context); (c) to conduct an experiment, examine how students access and make sense of different types of analytics, and validate their (potentially different) impact on behavioral, social, and conceptual engagement during the term and final grade for the discussion activity. The results can inform which areas of practice requires additional support for students' interpretable and actionable use of analytics. The research questions that guide my dissertation work are as follows:

RQ1. To what extent and how regularly do students access analytics?

RQ2. How do students make sense of the analytics?

RQ3. What impact does student use of analytics have on behavioral, social, and conceptual engagement in online collaborative reading?

RQ4. How do the different types of analytics affect students' access, sensemaking, and learning impact?

4 METHODOLOGY & CURRENT STATUS OF WORK

4.1 Phase 1: Problem analysis based on the synthesis of literature (completed)

As a first step, I reviewed the current status of knowledge from the studies examining student use of analytics in higher education. I identified the need for a framework to compare use practices across studies and developed a comprehensive framework of four different characteristics impacting student analytics use: context, information, tools/applications, and support for use. I then used this framing to examine and synthesize the documented practices of student analytics use (see Section 2.1). To probe how students learn to learn with data, specific attention was made to unpack the process of student analytic sensemaking by conceptualizing four components of analytic sensemaking drawn from the multidisciplinary fields (seeking information, decoding information, evaluating information, and making changes in behaviors or thoughts) and discussing research findings in relation to each of the four components (see Section 2.2). This phase led to the identification of existing gaps in the literature and design scope for future phases of work.

4.2 Phase 2: Designing intervention with different types of analytics (in process)

The second phase aims to design kinds of analytics and student practices of analytics use and based on the key lessons/issues identified in phase one: (a) four characteristics of analytics can heavily impact student access to and make sense of analytics and thus need to be considered for decisions in designing intervention; (b) especially, while support for student analytics use was widely provided in different forms of message design, their impact on learning was limitedly examined; (c) how student make sense of analytics can drive students to perceive different values of analytics and decide their use extent. This also can help students make change in their course engagement during the term and enhance better understanding of the course topics.

Based on these lessons, I am currently working on research design of phase two in designing research intervention that will offer students with two different versions of process-oriented analytics. Two courses that require students to participate in online collaborative reading activities will be chosen at a four-year private institution in the United States. The two kinds of analytics will be designed to show their progress in online collaborative reading activities to facilitate their behavioral, social, and conceptual engagement: (a) analytics with general description in a reporting form only with numbers (e.g. “This week your posts are mostly at a level one. To be at level two, please try to deeply reflect on the topics”) and (b) analytics with contextualized description that provides personalized examples from their own data (e.g. “Most of your posts are at a level one. Here's an example of one of your posts that was at a level one. And here's an example of one of your posts at the level two”). In Summer 2021, co-design work with a group of students who have taken those courses will be conducted to (a) decide desired qualities that the courses expect students to develop through collaborative reading activities and (b) design and develop different sets of analytic metrics that can explain these qualities. Then, the final version of the product from co-design will be used as an intervention in Fall 2021.

4.3 Phase 3: Implementing and examining student analytics use and sensemaking (planned)

In those courses, individual students will participate in online collaborative reading activities every week using the tool named Perusall throughout the semester. Intervention with the different types of the analytic reports will be assigned to each of the courses, and the analytics reports will be delivered through their emails on a weekly basis. During the term, students will be expected to use the analytic reports as part of their learning experiences by monitoring, reflecting on, and adapting their learning. During and after the semester, multiple data sources will be collected: one, student interview with walk-through of analytics to delve into the process of their analytic sensemaking, contributing factors, and learning changes they make in response to their interpretation; two, clickstream data from student access to the reports to investigate their patterns of analytics use; three, learning data from the online collaborative reading site (Perusall) to examine their patterns of analytics use and changes in behavioral, social, and cognitive engagement; four, post-surveys to identify their perceptions of analytics use, contributing factors, and areas for support they need for better analytics use. The findings from this work will be expected to contribute to understanding of (a) what part of analytics use students actually draw value from to enhance their learning, (b) how the support can prompt a particular form of analytic sensemaking, resulting in actual learning change, and (c) what kinds of learning support are needed to enhance student analytic sensemaking and actionable.

REFERENCES

Aguilar, S. J. (2018). Examining the relationship between comparative and self-focused academic data visualizations in at-risk college students' academic motivation. *Journal of Research on Technology in Education, 50*(1), 84-103.

Bennett, L., & Folley, S. (2019). Four design principles for learner dashboards that support student agency and empowerment. *Journal of Applied Research in Higher Education, 12*(1), 15-26.

Bodily, R., & Verbert, K. (2017, March). Trends and issues in student-facing learning analytics reporting systems research. In *Proceedings of the 7th International Conference on Learning Analytics & Knowledge* (pp. 309-318). New York, NY: ACM.

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.

Corrin, L., & De Barba, P. (2015, March). How do students interpret feedback delivered via dashboards?. In *Proceedings of the 5th International Conference on Learning Analytics & Knowledge* (pp. 430-431). New York, NY: ACM.

Foster, C., & Francis, P. (2020). A systematic review on the deployment and effectiveness of data analytics in higher education to improve student outcomes. *Assessment & Evaluation in Higher Education, 45*(6), 822-841.

Holman, C., Aguilar, S. J., Levick, A., Stern, J., Plummer, B., & Fishman, B. (2015, March). Planning for success: how students use a grade prediction tool to win their classes. In *Proceedings of the 5th International Conference on Learning Analytics & Knowledge* (pp. 260-264). New York, NY: ACM.

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017, September). Awareness is not enough: pitfalls of learning analytics dashboards in the educational practice. In *Proceedings of the ECTEL 2017* (pp. 82-96). Cham, Switzerland: Springer.

Jivet, I., Scheffel, M., Schmitz, M., Robbers, S., Specht, M., & Drachsler, H. (2020). From students with love: An empirical study on learner goals, self-regulated learning and sense-making of learning analytics in higher education. *The Internet and Higher Education, 47*, 100758.

Kia, F. S., Teasley, S. D., Hatala, M., Karabenick, S. A., & Kay, M. (2020, March). How patterns of students dashboard use are related to their achievement and self-regulatory engagement. In *Proceedings of the 10th International Conference on Learning Analytics & Knowledge* (pp. 340-349). New York, NY: ACM.

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.

Klein, G., Moon, B., & Hoffman, R. R. (2006). Making sense of sensemaking 1: Alternative perspectives. *IEEE Intelligent Systems, 21*(4), 70-73.

Lim, L. A., Dawson, S., Joksimovic, S., & Gašević, D. (2019b, March). Exploring students' sensemaking of learning analytics dashboards: Does frame of reference make a difference?. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (pp. 250-259). New York, NY: ACM.

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.

Pirolli, P., & Card, S. (2005, May). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis* (Vol. 5, pp. 2-4).

Prinsloo, P., & Slade, S. (2016). Student vulnerability, agency, and learning analytics: An exploration*. Journal of Learning Analytics, 3*(1), 159-182.

Teasley, S. D. (2017). Student facing dashboards: One size fits all?. *Technology, Knowledge and Learning, 22*(3), 377-384.

Wise, A. F., Zhao, Y., & Hausknecht, S. N. (2014). Learning analytics for online discussions: embedded and extracted approaches. *Journal of Learning Analytics, 1*(2), 48-71.

Using Computational Methods to Investigate Communicative Patterns in Educational Feedback

Jionghao Lin
Monash University
Jionghao.lin1@monash.edu

ABSTRACT: Many empirical studies in feedback research have been conducted to investigate effective practices and to measure the effects of feedback on learning processes and outcomes. However, there is much less research that looks at the quality of communication in the exchange of feedback between students and teachers. This doctoral research will focus on communicative patterns, mainly polite expressions and dialogue acts, in various types of feedback and aims to provide methods to convert feedback messages into polite communication forms to facilitate student learning in an online learning environment. Dialogue-based feedback will firstly be investigated with respect to the use of polite expressions and dialogue acts. Then, assessment-based feedback will be explored to analyze the association between communicative patterns and high-quality feedback. Finally, based on the findings derived from dialogue-based and assessment-based feedback, automatic methods will be proposed for converting feedback messages (e.g., from impolite to polite) to promote effective communication of feedback.

Keywords: Effective feedback, politeness communication, dialogue acts, learning analytics

1 BACKGROUND

Feedback about student learning has been considered as one of the most important factors in enhancing a student's academic achievement (Hattie & Timperley, 2007). In higher education, feedback is commonly used to provide information or comments on students' work to reveal a gap between the expected performance and the students' actual performance (Boud & Molloy, 2013). Based on feedback, students can recognize available options for achieving the expected learning goals and allocate more effort to minimize the gap or to pursue more challenging goals (Hattie & Timperley, 2007). It should be noted that feedback can assist students to achieve the expected learning goals only if the students recognize the usefulness of such feedback and make the corresponding adaptation of their learning strategies (Boud, 2015; Boud & Molloy, 2013). Therefore, students are not only the recipients of feedback but also the active agents to adapt their learning practices so as to achieve the expected outcome.

Feedback can be considered a communication process between a teacher and a student, which can be provided in an immediate or delayed manner (Hattie & Timperley, 2007). The immediate feedback is provided when students are in the process of completing their learning tasks (Hattie & Timperley, 2007). For example, feedback shared through a tutoring dialogue is a typical type of immediate feedback. In a tutorial dialogue, tutors can provide feedback after a student has asked a question or requested feedback. Feedback is considered a core component of tutorial dialogue (Jackson & Graesser, 2007), and the prior work found that dialogue feedback can motivate students to seek feedback and develop students' evaluative judgment (Carless, 2016). Whereas, delayed feedback is provided after students have submitted their learning products (Hattie & Timperley, 2007). For example, assignment feedback is a widely used type of delayed feedback. This type of feedback is released as a deliverable after students have submitted their assignments. Tutors evaluate students' learning performance according to relevant criteria or rubrics and provide feedback to assist the students' future learning. Both types of feedback are recognized to have positive effects for supporting students' learning tasks (Hattie & Timperley, 2007).

Feedback can be provided by educators, parents, students' peers, and computer-based systems tasks (Hattie & Timperley, 2007). Conventionally, in higher education, students receive feedback from educators in forms of rubrics and dialogue. A growing number of educational institutions provide an online learning environment which facilitates students' learning without the restrictions of fixed class capacity, timetables, and physical presence. Students can use online learning resources anytime and anywhere. However, many challenges exist in online settings regarding the provision of effective feedback to facilitate student learning. For example, the number of student enrollments (both online and on-campus) keeps increasing year by year but the number of teaching staff remains relatively steady, which results in a high student-teacher ratio. As a result, it becomes an unavoidable challenge for teachers to provide effective feedback to large student cohorts (L.-A. Lim et al., 2020; L.-A. Lim et al., 2019; Pardo, 2018; Pardo, Jovanovic, Dawson, Gašević, & Mirriahi, 2019).

1.1 Learning Analytics and Automated Feedback

Learning analytics (LA) offers much promise to address the issues of the feedback provision by providing salable and automated methods to deliver feedback for students (Pardo et al., 2019). An important topic in the LA field is the provision of automated, personalized and timely feedback to students (Pardo et al., 2019). LA development is accelerated by the increased adoption of learning technologies, which can automatically and unobtrusively collect data about students' learning activities. LA can utilize these data to enhance students' learning experiences based on evidence-based understanding and by providing personalized learning support (Greller & Drachsler, 2012; Pardo et al., 2019). For example, learning analytics dashboards (LADs) have received much attention in LA with their ability to provide students and teachers with automated feedback (Bodily & Verbert, 2017). However, the effectiveness of LADs for enhancing students' learning has been doubted by many researchers (Jivet, Scheffel, Specht, & Drachsler, 2018; L. Lim, Dawson, Joksimovic, & Gašević, 2019; Matcha, Gasevic, & Pardo, 2019). Several studies also emphasize that effective feedback relies on students' perception, understandings, and engagement (Price, Handley, & Millar, 2011; Winstone, Nash, Parker, & Rowntree, 2017), and current feedback based on LADs may not be consistent with the properties of effective feedback for assisting students' learning (L.-A. Lim et al., 2020; Matcha, Gasevic, et al., 2019). To offer effective feedback, it is vital to consider some human factors related to perception, cognition and engagement into the design of feedback.

It is worth noting that LA research already offers some software systems to provide personalized feedback, e.g., SRES (Liu, Bartimote-Aufflick, Pardo, & Bridgeman, 2017) and OnTask (Pardo et al., 2018). While these personalized feedback systems have demonstrated benefits in terms of improved learning and time management strategies (Matcha, Gašević, Uzir, Jovanović, & Pardo, 2019), learning outcomes (L.-A. Lim et al., 2019; Pardo et al., 2019), and satisfaction with feedback (Pardo et al., 2019), research that looks at the extent to which the way how feedback is worded can promote student motivation to engage with feedback is under-explored. To unlock the potential of feedback in students' learning process, the wording of feedback, such as the use of polite expression should be considered.

1.2 Communicative Patterns in Feedback

Given that feedback conveys evaluation about students' work, which can trigger different emotions with students, it is important to consider politeness, as a critical communicative dimension of effective feedback (Schneider, Nebel, Pradel, & Rey, 2015). According to Brown and Levinson (1987), politeness is centrally concerned with the way of treating people in communication, which takes into account other people's feelings. Since feedback is considered a form of communication, the role of politeness is worthy of investigation in educational feedback. Prior research found that expressing politeness in feedback can develop rapport and solidarity zone (Schallert et al., 2009), increase students' positive feelings, learning outcomes, and identity needs (Bolkan & Holmgren, 2012; Goodboy & Bolkan, 2009; Zhang & Sapp, 2013). Although some research found that providing students with polite feedback has benefits to a student's learning experience and outcomes, few studies investigated the role of politeness in other educational scenarios such as online learning.

In addition to politeness, other dimensions of communicative patterns should be considered in the analysis of feedback. For instance, when having a tutorial session, instructors and students typically formulate their utterances with specific intentions, e.g., requesting feedback and providing feedback. These utterances can be further interpreted as dialogue acts (e.g., **RF**-Request Feedback and **PF**-Provide Feedback). Generally, the communicative expressions between teachers and students can be formally interpreted as dialogue acts to indicate the intention hidden behind those expressions (Morrison & Rus, 2014). By comprehending dialogue acts, teachers and peers can better understand students and offer good advice. For example, when a student gives a correct answer, teachers can use praise to affirm the student's effort, which can motivate the student for further learning.

2 AIM OF THE RESEARCH

This doctoral research investigates the role of communicative patterns in feedback based on computational methods. The overall goal of this research is to contribute new insights into the role of communicative patterns (e.g., politeness and dialogue acts) in educational feedback. We will focus on feedback communicated in textual form through online learning environments. Specifically, we will investigate feedback offered in two types of settings: i) human-human tutorial dialogues offered as synchronous chats between students and tutors; and ii) assessment feedback provided on a piece of work by students and this feedback is typically communicated asynchronously. Therefore, we aim to answer the following research questions:

RQ.1 What is the role of politeness in providing feedback during text-based dialogic tutoring?

RQ.2 What is the role of dialogue acts in feedback exchanged in tutorial dialogue?

RQ.3 How are the communicative patterns, politeness and dialogue acts, associated with the quality of feedback in assessment feedback?

RQ.4 To what extent can we augment feedback with properties of effective communicative patterns in an online learning environment?

3 METHODOLOGY

This research will adopt computational methods (e.g., supervised machine learnings) to analyze the communicative patterns in educational feedback. There are three phases to answer the above RQs, which are shown in Figure 1. In phase one, we will investigate the politeness strategies and dialogue acts in tutorial dialogue to address RQ1 and RQ2, respectively. To answer RQ1, online tutoring dialogue data will be used to investigate the correlation between student performance and politeness features. The politeness features are composed by the politeness strategies detected by applying a politeness strategies identifier (Danescu-Niculescu-Mizil, Sudhof, Jurafsky, Leskovec, & Potts, 2013) and the level of politeness analyzed by politeness scoring methods (Niu & Bansal, 2018). To answer RQ2, the same tutorial dialogue data will be used to investigate the correlation between student performance and dialogue acts. A subset of tutorial dialogues will be manually coded and the annotated data will be used for training an automated dialogue acts classifier for analyzing the whole tutorial dataset. In phase two, we will answer RQ3 by using the dialogue acts classifier developed in Phase 1 and the politeness tools (Danescu-Niculescu-Mizil et al., 2013; Niu & Bansal, 2018). The dialogue acts classifiers and politeness tools are adopted to investigate the correlation between communication patterns (politeness and dialogue acts) and effective assessment feedback. In Phase 3, we will build a politeness feedback communication convertor based on the findings from Phase 1 and Phase 2 to address RQ4. More specifically, a method for automatic transformation of text into polite expressions (Madaan et al., 2020) will be used as a foundation for the development of our method for enhancing politeness expressions in feedback.

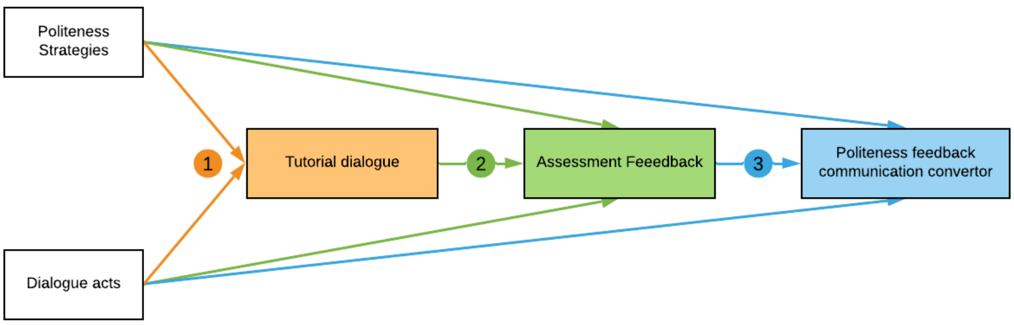


Figure 1: Research Phases

4 CURRENT PROGRESS

In this research, the tutorial dataset was provided by an online tutoring service Yup.com and the assessment feedback by a Brazilian public university. We have obtained the ethical approvals by the Monash University Human Research Ethics Committee to conduct research on these two datasets. In Year 1 of this project, RQ1 has been addressed, and a short paper was published in the Proceedings of the 21st International Conference of Artificial Intelligence in Education. In this paper (Lin, Lang, Xie, Gašević, & Chen, 2020), we adopted the politeness strategies identifier to extract the politeness strategies in the tutorial dialogue. Additionally, we defined a metric called **UP-score** (**P**oliteness **score** of an **U**tterance) to measure the level of politeness for each utterance from students and tutors. According to the work done by Danescu-Niculescu-Mizil et al. (2013), the polite strategies were assigned with positive numeric values, and impolite strategies were assigned with negative numeric values. Then, the **UP-score** for each utterance is obtained by subtracting the number of polite strategies with the number of impolite strategies (see Equation 1).

$$\textbf{UP-Score} = \#\ \text{Polite strategies} - \#\ \text{Impolite strategies} \quad (1)$$

A positive UP-Score implies that the utterance is polite, while a negative value suggests an impolite one. The study results present that: i) tutors in sessions where students successfully complete the learning tasks were likely to use impolite/direct expression (e.g., "Solve it for getting the value of x")to guide students after the first few utterances, and ii) students who successfully solved tasks were slightly more polite than those without at the beginning of a tutorial session.

Additionally, we conducted a further study of investigating the role of politeness in tutorial dialogue, and the results are being prepared for journal submission. In this study, we adopted the politeness scoring method to extract the politeness scores for each utterance in the tutorial dialogue, and used the politeness strategy identifier to identify the politeness used in the tutorial dialogue. We found that: i) the trending results of the politeness level extracted by politeness scoring tool are consistent with the results measured by UP-score, ii) the students who successfully solved tasks tended to use more polite expressions at the beginning of a tutorial session, iii) the positive correlation between the use of polite expressions and students' performance was more observed in the group of students who had made good task progress vs. those who had not, iv) politeness alone might not be sufficient to predict student performance and other factors should also be taken into account (e.g., the amount of time that a student spent in solving a task).

REFERENCES

Bodily, R., & Verbert, K. (2017). *Trends and issues in student-facing learning analytics reporting systems research.* Paper presented at the Proceedings of the seventh international learning analytics & knowledge conference.

Bolkan, S., & Holmgren, J. L. (2012). "You are such a great teacher and I hate to bother you but…": Instructors' perceptions of students and their use of email messages with varying politeness strategies. *Communication Education, 61*(3), 253-270.

Boud, D. (2015). Feedback: ensuring that it leads to enhanced learning. *Clinical Teacher*.

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in higher education, 38*(6), 698-712.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (Vol. 4): Cambridge university press.

Carless, D. (2016). Feedback as dialogue. *Encyclopedia of educational philosophy and theory*, 1-6.

Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., & Potts, C. (2013). A computational approach to politeness with application to social factors. *arXiv preprint arXiv:1306.6078*.

Goodboy, A. K., & Bolkan, S. (2009). College teacher misbehaviors: Direct and indirect effects on student communication behavior and traditional learning outcomes. *Western Journal of Communication, 73*(2), 204-219.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society, 15*(3), 42-57.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research, 77*(1), 81-112.

Jackson, G. T., & Graesser, A. C. (2007). Content matters: An investigation of feedback categories within an ITS. *Frontiers in Artificial Intelligence and Applications, 158*, 127.

Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018). *License to evaluate: Preparing learning analytics dashboards for educational practice.* Paper presented at the Proceedings of the 8th International Conference on Learning Analytics and Knowledge.

Lim, L., Dawson, S., Joksimovic, S., & Gašević, D. (2019). *Exploring students' sensemaking of learning analytics dashboards: Does frame of reference make a difference?* Paper presented at the Proceedings of the 9th International Conference on Learning Analytics & Knowledge.

Lim, L.-A., Dawson, S., Gašević, D., Joksimović, S., Pardo, A., Fudge, A., & Gentili, S. (2020). Students' perceptions of, and emotional responses to, personalised learning analytics-based feedback: an exploratory study of four courses. *Assessment & Evaluation in higher education*, 1-21.

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*, 101202.

Lin, J., Lang, D., Xie, H., Gašević, D., & Chen, G. (2020). *Investigating the Role of Politeness in Human-Human Online Tutoring.* Paper presented at the International Conference on Artificial Intelligence in Education.

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: Fundaments, applications, and trends* (pp. 143-169): Springer.

Madaan, A., Setlur, A., Parekh, T., Poczos, B., Neubig, G., Yang, Y., . . . Prabhumoye, S. (2020). Politeness transfer: A tag and generate approach. *arXiv preprint arXiv:2004.14257*.

Matcha, W., 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*.

Matcha, W., Gašević, D., Uzir, N. A. A., Jovanović, J., & Pardo, A. (2019). *Analytics of learning strategies: Associations with academic performance and feedback.* Paper presented at the Proceedings of the 9th International Conference on Learning Analytics & Knowledge.

Morrison, D. M., & Rus, V. (2014). –Moves, Tactics, Strategies, and Metastrategies: Defining the Nature of Human Pedagogical Interaction. *Design recommendations for intelligent tutoring systems*, 217.

Niu, T., & Bansal, M. (2018). Polite dialogue generation without parallel data. *Transactions of the Association for Computational Linguistics, 6*, 373-389.

Pardo, A. (2018). A feedback model for data-rich learning experiences. *Assessment & Evaluation in higher education, 43*(3), 428-438.

Pardo, A., Bartimote, K., Shum, S. B., Dawson, S., Gao, J., Gašević, D., . . . Mirriahi, N. (2018). OnTask: Delivering data-informed, personalized learning support actions. *Journal of Learning Analytics, 5*(3), 235–249-235–249.

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.

Price, M., Handley, K., & Millar, J. (2011). Feedback: Focusing attention on engagement. *Studies in higher education, 36*(8), 879-896.

Schallert, D. L., Chiang, Y.-h. V., Park, Y., Jordan, M. E., Lee, H., Cheng, A.-C. J., . . . Song, K. (2009). Being polite while fulfilling different discourse functions in online classroom discussions. *Computers & Education, 53*(3), 713-725.

Schneider, S., Nebel, S., Pradel, S., & Rey, G. D. (2015). Mind your Ps and Qs! How polite instructions affect learning with multimedia. *Computers in Human Behavior, 51*, 546-555.

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.

Zhang, Q., & Sapp, D. A. (2013). Psychological reactance and resistance intention in the classroom: Effects of perceived request politeness and legitimacy, relationship distance, and teacher credibility. *Communication Education, 62*(1), 1-25.

Towards clinical practice analytics: visualising repurposed routinely collected clinical indicator data to support reflection

Bernard Bucalon
The University of Sydney
bernard.bucalon@sydney.edu.au

ABSTRACT: This presentation will outline a PhD thesis that aims to create personalised dashboards based on hospital clinical indicator data to enable clinicians to reflect on their practice - the work is part of a new field we call "Clinical Practice Analytics". The need for new data tools to support clinical practice is driven by medical regulators prioritising the completion of professional development activities that review a clinician's actual performance and measure patient health outcomes. Despite the known reliability and validity issues with clinical indicators, there is interest from health professionals to use routinely collected patient administrative data for performance feedback and personalised professional development. A systematic scoping review is in progress to map the literature on existing state-of-the-art clinical dashboards that support reflection and life-long learning. Ethics approval has been received from a Sydney-based hospital ethics committee to conduct a user study with clinicians to co-design and evaluate a practice analytics dashboard prototype. The primary contribution of this research will be the development of a novel dashboard prototype, informed by a long-term multi-site field study, that will enable clinicians to reflect on their personal performance data. The research findings will also contribute to our understanding of the new field of Clinical Practice Analytics.

Keywords: learning analytics dashboards, clinical practice analytics, visualisation, reflection

1 BACKGROUND

This thesis[1] is part of the Digital Health Cooperative Research Centre (DHCRC) Practice Analytics program. The multi-disciplinary research team from the fields of medical anthropology, psychology, data science, and human-centred technology are exploring the new field of "Clinical Practice Analytics", which investigates how data extracted from sources such as Patient Administration Systems can be re-purposed by individual practitioners and clinical teams to engage in reflective practice, reduce variation, and support performance improvement.

Medical regulatory bodies in Australasia, through the *Professional Practice Framework*, have shifted the emphasis of mandatory professional development from traditional educational activities such as courses and conferences, to activities that review a clinician's actual performance (multi-source feedback) and that measure patient health outcomes (clinical audit & feedback) (MBA, 2017). To meet this need, hospitals have explored designing dashboards for clinicians to access clinical indicators, based on routinely collected administrative data (length of stay, readmissions, hospital acquired complications) to provide valuable insight into their personal clinical performance.

[1] This PhD project is funded by the Digital Health Cooperative Research Centre (DHCRC).

Learning analytics theory and methods have been applied to medical education to gain insights about learner- or system-level performance. Studies have primarily focused on the assessment of medical students and residents in specialist training programs, such as programmatic assessment of Emergency Medicine residents, online modules to teach x-ray interpretation, and the design of analytics dashboards for Internal Medicine (Chan et al., 2018). While Board examinations and work-based assessments certify initial competence of medical graduates, practicing clinicians require ongoing self-assessment to maintain standards and to identify improvement needs.

Dashboards designed to improve the sense-making and decision-making around a clinician's own performance should support self-regulated learning activities (Matcha et al., 2020), as reflective practice and life-long learning underpin the continuing professional development (CPD) frameworks mandated by the Medical Board of Australia (MBA, 2018). This thesis will leverage learning analytics techniques on re-purposed clinical indicator data to support reflection and performance improvement, therefore contributing to our understanding of Clinical Practice Analytics.

2 CURRENT KNOWLEDGE OF PROBLEM DOMAIN

Despite the known organisational, cultural, and technical issues with collecting and reporting clinical indicators (Freeman, 2002), attitudes of health professionals indicate there is an appetite for easier and timely access to routinely clinical indicator data for performance feedback (Shaw et al., 2019). There is an established body of literature investigating the use of visualisation dashboards in clinical settings to improve accuracy (decision-making), efficiency (task completion time), usability (user satisfaction), and quality & safety (adherence to guidelines) (Brown et al., 2018; Faiola et al., 2015; Gude et al., 2017). However, the studies generally are based on formal "audit and feedback" processes, rather than formative professional development, which have shown limited effectiveness in changing behaviour (Ivers et al., 2012).

Dashboards can be improved through designing "targeted reflection and learning planning activities based on information in the dashboard" (Boscardin et al., 2018). Learning analytics dashboards can support metacognitive processes such as self-monitoring, reflection, and planning (Schwendimann et al., 2017). While few dashboard evaluations take into account the educational concepts that were used as theoretical foundations for their design (Jivet et al., 2018), learner modelling visualisations - called open learner models (OLMs) - can improve the accuracy of underlying learner models by taking into account the learners' viewpoints on their learning, while also promoting learner reflection (Bull et al., 2016).

The design of learning analytics dashboards for reflection and performance improvement must address the following challenges: existing organisational and cultural barriers to performance feedback (Pooley et al., 2019), accounting for the validity and reliability of clinical indicator data e.g. co-morbidities as a contributing factor (Mainz, 2003), establishing a learner model to underpin the data visualisations (Bodily et al., 2018), and the complexity of designing visualisations for individuals verses teams whilst communicating uncertainty (Shneiderman et al., 2013).

3 RESEARCH QUESTIONS

While there have been studies into the use of dashboards within clinical settings (Dowding et al., 2015; Khairat et al., 2018), there are few studies that investigate the visualisation of clinical indicator data for clinicians within the context of continuing professional development, particularly for reflective practice and life-long learning (Fox et al., 2016; Janssen et al., 2020; Martinez et al., 2018).

The thesis aims to answer the following research questions:

- What are the key design opportunities and trade-offs for clinical practice analytics dashboards?
- How do clinicians make use of hospital clinical indicators to gain insights about their personal and team practice?
- How can we create a personalised dashboard that can enable clinicians to gain insights about their own practice?

4 PROPOSED SOLUTION

Through collaborations with the DHCRC research partner hospital sites, this thesis plans to address the research questions by engaging with clinicians in a co-design process to understand their priority clinical indicators, visualise the re-purposed clinical indicators; to ultimately design and evaluate a personalised learning analytics dashboard to support reflective practice.

We envision the learning analytics dashboard will be personalised, tailored to the clinician's scope of practice and career stage. The dashboard should support the learner to recall the data they have and have not seen, compare their performance over suitable time-periods (6-12 months) to view changes in practice, and adjust for changes in team membership when colleagues move to different hospitals. The dashboard should also incorporate interface scaffolding to help clinicians consider questions about their own performance for effective reflection. For example, scaffolding techniques such as tutorials and goal prompts were used to promote reflection of personal physical tracking data (Tang & Kay, 2018).

The project will focus on medical specialists such as physicians, paediatricians, surgeons, radiologists, anaesthetists, obstetricians & gynaecologists, emergency and intensive care specialists. We may design a specialty specific dashboard should the opportunity arise. Clinical indicators will be primarily sourced from medical, surgical, and patient administrative databases collected within private hospitals. Examples of clinical indicators include length of stay, re-admissions, hospital acquired complications (HACs), unplanned returns to theatre, and late discharges. The project will not focus on medical students, post-graduate trainees in specialist training programs (residents/registrars), general practitioners (GPs), nurses, or allied health practitioners. Electronic Medical Record (EMR) data will be excluded e.g., patient history, vitals, medications, and laboratory results. While the project is focused on practice-level data rather than individual patient-level data, the dataset provided by the research partners does include International Classification of Diseases (ICD-10) and Medicare Benefits Schedule (MBS) codes which reveals more detailed information about symptoms, diseases, and treatments.

5 METHODOLOGY

A **scoping review** is in the manuscript writing stage. The review aimed to map existing literature on dashboards and visualisation studies of re-purposed hospital clinical indicator data to support reflective practice. The review found varied approaches to the design, implementation, and evaluation of such interventions. Interface elements to support self-regulated learning such as reflection, goal setting, and strategic planning were sparse.

A **user study** at a private hospital based in Sydney will recruit 30 clinicians to participate in semi-structured interviews and focus groups. Thematic analysis of the qualitative data collected will identify the design requirements for the dashboard visualisations. A dashboard prototype will be developed using coded real-world hospital administrative data from 2018 and 2019. A Think-Aloud Protocol (TAP) and standard usability questionnaires will be used during user interviews with clinicians to gather rich feedback (Bangor et al., 2008; Sauro & Dumas, 2009).

The findings from the single-site user study will guide the design of **a multi-site study** across private hospital sites in Sydney, Melbourne, and Perth. **Phase 1** plans to recruit at least 15 clinicians at each study site, representing two to three different medical specialties. The dashboard from the previous need finding study will be iterated on and then evaluated in laboratory settings to assess pure usability and its effectiveness in enabling clinicians to identify important features of their practice.

In **Phase 2**, the prototype will be iterated on and clinicians will be asked to reflect on their personal performance data using the prototype dashboard in authentic settings, over a 6-month period. Field work including direct observation, interviews, and focus groups will be used to evaluate the impact of the prototype dashboard on clinician reflection and long-term learning. Field observations will provide rich qualitative data into how the dashboard is used in practice to address any problems with implementation and adoption (Buckingham Shum et al., 2019). Data analysis on the system logs will be conducted to understand usage patterns. Data analysis will be conducted on the underlying clinical indicators in the hospital administrative database to identify changes in practice.

6 CURRENT STATUS OF THE WORK

The thesis commenced in March 2020 and the key activities completed to date: drafted an initial literature review and research plan; completed scoping review screening and analysis; received ethics approval for first user study and commenced participant recruitment; presented research project at local Human-Centred Technology research seminar, OzCHI Doctoral Consortium (Dec, 2020), and Digital Health Week (Feb, 2021).

REFERENCES

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human–Computer Interaction*, *24*(6), 574–594. https://doi.org/10.1080/10447310802205776

Bloom, B. S. (2005). Effects of continuing medical education on improving physician clinical care and patient health: A review of systematic reviews. *International Journal of Technology Assessment in Health Care*, *21*(3), 380–385. https://doi.org/10.1017/S026646230505049X

Bodily, R., Kay, J., Aleven, V., Jivet, I., Davis, D., Xhakaj, F., & Verbert, K. (2018). Open learner models and learning analytics dashboards: A systematic review. *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, 41–50. https://doi.org/10.1145/3170358.3170409

Boscardin, C., Fergus, K. B., Hellevig, B., & Hauer, K. E. (2018). Twelve tips to promote successful development of a learner performance dashboard within a medical education program. *Medical Teacher*, *40*(8), 855–861. https://doi.org/10.1080/0142159X.2017.1396306

Brown, B., Balatsoukas, P., Williams, R., Sperrin, M., & Buchan, I. (2018). Multi-method laboratory user evaluation of an actionable clinical performance information system: Implications for usability and patient safety. *Journal of Biomedical Informatics*, *77*, 62–80. https://doi.org/10.1016/j.jbi.2017.11.008

Buckingham Shum, S., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-Centred Learning Analytics. *Journal of Learning Analytics*, *6*(2). https://doi.org/10.18608/jla.2019.62.1

Bull, S., Ginon, B., Boscolo, C., & Johnson, M. (2016). Introduction of learning visualisations and metacognitive support in a persuadable open learner model. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*, 30–39. https://doi.org/10.1145/2883851.2883853

Cervero, R. M., & Gaines, J. K. (2015). The Impact of CME on Physician Performance and Patient Health Outcomes: An Updated Synthesis of Systematic Reviews: *Journal of Continuing Education in the Health Professions*, *35*(2), 131–138. https://doi.org/10.1002/chp.21290

Chan, T., Sebok-Syer, S., Thoma, B., Wise, A., Sherbino, J., & Pusic, M. (2018). Learning Analytics in Medical Education Assessment: The Past, the Present, and the Future. *AEM Education and Training*, *2*(2), 178–187. https://doi.org/10.1002/aet2.10087

Dowding, D., Randell, R., Gardner, P., Fitzpatrick, G., Dykes, P., Favela, J., Hamer, S., Whitewood-Moores, Z., Hardiker, N., Borycki, E., & Currie, L. (2015). Dashboards for improving patient care: Review of the literature. *International Journal of Medical Informatics*, *84*(2), 87–100. https://doi.org/10.1016/j.ijmedinf.2014.10.001

Faiola, A., Srinivas, P., & Duke, J. (2015). Supporting Clinical Cognition: A Human-Centered Approach to a Novel ICU Information Visualization Dashboard. *AMIA Annual Symposium Proceedings*, *2015*, 10.

Fox, L. A., Walsh, K. E., & Schainker, E. G. (2016). The Creation of a Pediatric Hospital Medicine Dashboard: Performance Assessment for Improvement. *Hospital Pediatrics*, *6*(7), 412–419. https://doi.org/10.1542/hpeds.2015-0222

Freeman, T. (2002). Using performance indicators to improve health care quality in the public sector: A review of the literature. *Health Services Management Research*, *15*(2), 126–137. https://doi.org/10.1258/0951484021912897

Gude, W. T., van Engen-Verheul, M. M., van der Veer, S. N., de Keizer, N. F., & Peek, N. (2017). How does audit and feedback influence intentions of health professionals to improve practice? A laboratory experiment and field study in cardiac rehabilitation. *BMJ Quality & Safety*, *26*(4), 279–287. https://doi.org/10.1136/bmjqs-2015-004795

Ivers, N., Jamtvedt, G., Flottorp, S., Young, J. M., Odgaard-Jensen, J., French, S. D., O'Brien, M. A., Johansen, M., Grimshaw, J., & Oxman, A. D. (2012). Audit and feedback: Effects on professional practice and healthcare outcomes. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD000259.pub3

Janssen, A., Donnelly, C., Kay, J., Thiem, P., Saavedra, A., Pathmanathan, N., Elder, E., Dinh, P., Kabir, M., Jackson, K., Harnett, P., & Shaw, T. (2020). Developing an Intranet-Based Lymphedema Dashboard for Breast Cancer Multidisciplinary Teams: Design Research Study. *Journal of Medical Internet Research*, *22*(4), e13188. https://doi.org/10.2196/13188

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*, 31–40. https://doi.org/10.1145/3170358.3170421

Khairat, S. S., Dukkipati, A., Lauria, H. A., Bice, T., Travers, D., & Carson, S. S. (2018). The Impact of Visualization Dashboards on Quality of Care and Clinician Satisfaction: Integrative Literature Review. *JMIR Human Factors*, *5*(2), e22. https://doi.org/10.2196/humanfactors.9328

Mainz, J. (2003). Defining and classifying clinical indicators for quality improvement. *International Journal for Quality in Health Care*, *15*(6), 523–530. https://doi.org/10.1093/intqhc/mzg081

Martinez, D. A., Kane, E. M., Jalalpour, M., Scheulen, J., Rupani, H., Toteja, R., Barbara, C., Bush, B., & Levin, S. R. (2018). An Electronic Dashboard to Monitor Patient Flow at the Johns Hopkins Hospital: Communication of Key Performance Indicators Using the Donabedian Model. *Journal of Medical Systems*, *42*(8), 133. https://doi.org/10.1007/s10916-018-0988-4

Matcha, W., Uzir, N. A., Gasevic, D., & Pardo, A. (2020). A Systematic Review of Empirical Studies on Learning Analytics Dashboards: A Self-Regulated Learning Perspective. *IEEE Transactions on Learning Technologies*, *13*(2), 226–245. https://doi.org/10.1109/TLT.2019.2916802

MBA. (2017). *Expert Advisory Group on Revalidation Final Report*. Medical Board of Australia. http://www.medicalboard.gov.au/documents/default.aspx?record=WD17%2F24295&dbid=AP&chksum=Txmn8C7v%2BC53Wjsz3sXn2w%3D%3D

MBA. (2018). *Professional Performance Framework*. Medical Board of Australia. https://www.medicalboard.gov.au/Registration/Professional-Performance-Framework.aspx

Pooley, M., Pizzuti, C., & Daly, M. (2019). Optimizing Multisource Feedback Implementation for Australasian Physicians: *Journal of Continuing Education in the Health Professions*, *39*(4), 228–235. https://doi.org/10.1097/CEH.0000000000000267

Sauro, J., & Dumas, J. (2009). Comparison of three one-question, post-task usability questionnaires. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1599–1608.

Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., & 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

Shaw, T., Janssen, A., Crampton, R., O'Leary, F., Hoyle, P., Jones, A., Shetty, A., Gunja, N., Ritchie, A. G., Spallek, H., Solman, A., Kay, J., Makeham, M. A., & Harnett, P. (2019). Attitudes of health professionals to using routinely collected clinical data for performance feedback and personalised professional development. *Medical Journal of Australia*, *210*(S6). https://doi.org/10.5694/mja2.50022

Shneiderman, B., Plaisant, C., & Hesse, B. W. (2013). Improving Healthcare with Interactive Visualization. *Computer*, *46*(5), 58–66. https://doi.org/10.1109/MC.2013.38

Tang, L. M., & Kay, J. (2018). Scaffolding for an OLM for Long-Term Physical Activity Goals. *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, 147–156. https://doi.org/10.1145/3209219.3209220

Building theory-informed learning analytics to understand and intervene in Socially-Shared Regulation of Learning

Cristina Villa-Torrano
GSIC-EMIC Research Group, Universidad de Valladolid, Spain
cristina@gsic.uva.es

ABSTRACT: There is a need to strengthen regulatory processes in collaborative learning. The Socially-Shared Regulation of Learning (SSRL) theory aims at understanding the regulatory processes through which group members negotiate objectives, planning, and strategies for carrying out a collaborative activity. Some studies on this topic have been conducted using students' self-reported or physiological data. However, self-reported data is biased by the students' perception and invasive sensors are costly and cumbersome. Moreover, these studies do not provide actionable information on time. Additionally, the analysis of SSRL becomes even more challenging when not restricted to specific learning environments or learning situations. Therefore, we propose to use Learning Design to guide data collection and inform the learning analytics using trace data from different technological tools. Through this, we expect to build predictive models that provide actionable information on SSRL, with a methodology that is not restricted to a specific learning design.

Keywords: Socially-Shared Regulation of Learning, Collaborative Learning, Learning Analytics, Learning Design.

1 BACKGROUND

Collaboration is one of the 21st Century Skills (Voogt & Roblin, 2010) that is increasingly present in academic and work context (Malmberg et al., 2015). Collaborating with others benefits learning yet comes with some challenges (Kreijns et al., 2003) that students need to overcome with their peers to achieve the shared learning goals (Malmberg et al., 2015). As noted by (Järvelä et al., 2020), success in collaborative learning often occurs when team members systematically activate and maintain their cognition, motivation, and emotions towards the achievement of their shared goals, i.e., socially regulating team efforts. Moreover, many empirical studies show that regulatory processes are critical for the success of collaborative learning (Järvelä et al., 2016).

Socially-Shared Regulation of Learning (SSRL) is a field in the framework of self-regulated learning theories that integrates different types of collective regulatory processes that contribute to shared regulation (Hadwin et al., 2011). Shared regulation processes happen when team members negotiate the perception of tasks, objectives, planning, and strategies. SSRL is theorized to consist of four stages that are interconnected and can be recursive (Malmberg et al., 2015): i) negotiation and construction of the perception of the task, based on internal and external representations; ii) sharing of objectives and generating plans to achieve them; iii) coordination and monitoring of progress; iv) reflection and redesign of objectives, planning or perception of activities. There exists initial evidence that successful groups are those that use multiple regulatory processes; students start using self-regulatory processes, such as task understanding and monitoring, and then perform shared regulation processes, such as jointly making plans for how to approach the task (Malmberg et al., 2015).

SSRL has already been explored from several perspectives. There are works where SSRL is studied using self-reported data about the challenges perceived by the groups and analyzing what SSRL strategies they develop to overcome them (Malmberg et al., 2015). In other works, such as (Malmberg et al., 2017), groups collaboratively carry out an assignment and then have to answer a questionnaire related to shared understanding, challenges, planning, etc. In this case, SSRL is studied through the conversations that students have through an online platform. Recently, SSRL has been researched by analyzing physiological data, observation data (video) and expression recognition (Järvelä et al., 2019). Three main limitations can be identified in these works: i) the validity of their findings is limited to very specific platforms and learning situations, and they do not consider how the pedagogical design and intentions shape the collaborative behavior and relevant regulatory processes; ii) the data was obtained through self-reported instruments or invasive sensors. However, the literature shows that students are biased when asked what regulatory processes they have followed (Saint et al., 2020), while the use of invasive sensors is less likely to be widely accepted; and iii) the focus of these studies has been on understanding regulatory processes post-hoc, but not in supporting these regulatory processes with actionable information for teachers and/or students during the enactment of the learning situation.

Previous work has shown that one way of including contextual and pedagogical information in the analysis is by means of the Learning Design (LD) (Rodríguez-Triana et al., 2015). Over the last two decades, the LD research field has been proposing processes and tools aimed at effectively supporting the complex task of conceptualizing and elaborating activity plans that can be enacted, shared and repurposed (Conole, 2013; Mor & Craft, 2012). Previous works suggest that LD can help in collecting learning data, in making meaning out of it, and in analyzing it (Lockyer et al., 2013). Therefore, and if we do not want to propose ways of analyzing SSRL processes that are restricted to concrete and specific learning situations, we need to propose LA approaches for SSRL that can be applied to learning environments in which different learning designs can be supported and enacted. In such LA approaches to SSRL, LD would be expected to play a significant role. Virtual Learning Environments, Distributed Learning Environments, and even MOOC platforms are examples of such environments that can support different learning situations and that, at the same time, provide trace data about students' behavior. Using this trace data, instead of (more biased) self-reported data and/or (difficult to collect) physiological data, we expect, on the one hand, to be able to understand shared regulation processes and, on the other hand, to be able to detect optimal and sub-optimal patterns of shared regulation during the different phases that are expected to happen, according to the learning design, to be able to make early interventions. For the latter, we expect to build predictive models based on the optimal and sub-optimal processes detected in order to provide actionable indicators.

Some of these features are being pursued by current research in the area of self-regulated learning (Jovanovic et al., 2020; Saint et al., 2020), where the traces of online tools are coded into macro-level constructs (e.g., planning) which comprise micro-level actions (e.g., setting goals or making personal plans) based on the theoretical models of SRL. In these studies, researchers detect predictable patterns that could inform the development of automatic interventions to provide real-time feedback (Saint et al., 2020). In this thesis, we follow a similar approach where the learning designs would help identifying macro-level constructs identified by SSRL theories and the micro-level actions would correspond to the students' actions. We expect to detect optimal and sub-optimal patterns of shared regulation from the traces of online platforms and that these patterns can help to create predictive models with actionable information. To the best of our knowledge, this approach has not been followed in the SSRL literature.

2 RESEARCH QUESTIONS AND GOALS

The underlying research question of this doctoral thesis is: **How can Learning Analytics based on the theory of SSRL help identify and predict patterns of shared regulation that provide actionable information in collaborative learning situations using trace data?** Our approach to answer this question is to automatically extract meaningful features from trace data considering the learning design that defines the collaborative learning situation in which SSRL processes are expected to happen. The general objective (to provide actionable information by detecting patterns of SSRL using trace data) is divided into two particular objectives:

1. **To map event data to SSRL theory constructs.**
   According to (Siadaty et al., 2016), precise conceptual SSRL models need to be defined. Based on a specification of SSRL constructs, we will explore how teachers can be involved to inform learning designs with additional information about where and when regulatory processes are expected to happen. This will help match trace data produced in the enactment of the activities to the appropriate SSRL construct. The mapping of traces to SSRL phases/constructs can help to identify important features to detect shared regulation patterns.

2. **To provide actionable information by building early predictive models of successful collaboration based on SSRL patterns.**
   It should be explored which learning analytics techniques could detect optimal and sub-optimal SSRL patterns through the mapped data. Once the above objective is achieved, it will be possible to identify which features can help to make early predictions. Furthermore, the detection of optimal and sub-optimal SSRL patterns would provide actionable information for early interventions.

3 BRIEF STATE OF THE ART

In recent years, a number of empirical studies have been conducted in the area of SSRL. In particular, a learning environment with regulation tools was used in (Malmberg et al., 2015) to prompt students to recognize challenges that may hinder collaboration and to develop SSRL strategies to overcome them. This study employs students' self-reported answers to the questions asked in the virtual environment, coded by the authors. The result of this research indicates that there is a difference between the regulatory processes followed by high and low performing groups. On the other hand, (Malmberg et al., 2017) focuses on the temporal and sequential order of the different types of regulation (self-regulation, co-regulation and socially shared regulation of learning) in collaborative activities. The data used in the study consists of videos of the working groups during two months in a math didactics course. Finally, in (Järvelä et al., 2019), a preliminary study uses data from different sources to help understand SSRL processes. Specifically, the use of physiological sensors is explored in greater depth, as is also detailed in (Järvelä et al., 2020).

These studies have been carried out with self-reported data or physiological data from students using invasive sensors. However, regulation can also be mapped to dynamic series of events that change over the learning situation (Siadaty et al., 2016) using traces from learning platforms. Furthermore, the studies mentioned on SSRL focus on understanding the processes of shared regulation, but not on making early predictions that allow timely interventions. This approach has recently started to be researched in the area of Self-Regulated Learning (SRL) through process mining (Saint et al., 2020). These works detect predictable patterns that could provide actionable information to trigger feedback in real time. However, to the best of our knowledge, it has not been researched in the area of SSRL. Moreover, these studies do not consider the context and pedagogical intentions behind each activity of a learning situation. As we mentioned before, regulation and social processes change along the learning situation (Malmberg et al., 2015). Therefore, it is important to align the learning design and learning analytics in order to: i) inform about the processes that are expected to occur during the situation (Er et al., 2019); and ii) guide the collection of data and the analysis to be made. Although the connection between learning design and learning analytics is growing significantly in the literature (Lockyer et al., 2013; Rodríguez-Triana et al., 2015), to the best of our knowledge, it is not being considered in the area of SSRL.

4 METHODOLOGY

The proposed methodology to answer the research question is Design Science Research Methodology (DSRM) (Peffers et al., 2007). DSRM aims at the creation and evaluation of artifacts that solve problems, like constructs, models or any designed object that offers a solution to the research problem. This methodology defines a process model involving the following phases: (i) identify a problem and motivate its interest; (ii) define the objectives of a solution; (iii) design and develop an artifact for the solution; (iv) demonstrate how the artifact solves the problem; (v) evaluate it; and (vi) communicate its performance. These phases do not need to happen necessarily sequentially. Indeed, refinements of the proposed solutions are foreseen by iteration through the different activities.

The overarching objectives of this thesis and its iterative nature make DSRM a suitable methodology to frame this thesis work. This PhD thesis aims to design and develop artifacts that provide actionable information by detecting patterns of SSRL using trace data. During the thesis, we need to involve the main stakeholders (teachers, learning/instructional designers, students, …) with several purposes, including: identify and describe learning scenarios that can benefit from SSRL, explore how teachers can be involved to inform learning designs with additional information about where and when regulatory processes are expected to happen, evaluate the degree in which the solutions meet the needs of the participants, etc.

Regarding the number of iterations needed, we foresee three iterations. The first iteration consists of a literature review focusing on theoretical models and the adoption of these models in empirical studies to support collaboration. This literature review is complemented with an exploration of the relevant data sources, machine learning techniques and actionable information to generate in relation to SSRL. Moreover, a first conceptual solution is proposed, and it is evaluated by exploratory studies, that will help in turn to understand better the problem and the goals. During the second iteration, the conceptual and technological solution to solve the detected gaps will be refined and developed. It is expected that, during this iteration, two studies can be conducted to evaluate part of the proposal. Finally, in the third iteration the proposal will be improved with the previous evaluations and the final evaluation of the proposed solution will be carried out.

5 CURRENT PROGRESS

So far, the author has been working on the first iteration of the thesis plan. She has carried out a non-systematic review of the state of the art of SSRL and SRL, focusing on the definition of the theoretical models, the adoption of these models in empirical studies and the types of data collected. In addition, the author, and her colleagues have submitted a paper to an international conference where they work in an exploratory collaborative scenario where they detect shared regulation processes through trace data. The collected data was coded based on the theoretical model and they detect SSRL processes using a process mining technique. The theory-informed LA also helped to interpret the processes of shared regulation and to detect behavior that was not expected during the activity. However, this study was conducted using data coming from an online learning platform designed to support a specific type of collaborative activity and the learning design was very concrete. Furthermore, this exploratory scenario has helped us to identify different aspects of a specific collaborative learning scenario: regulation processes that occur, data that can be collected, interventions that can be made, … As a result, it will facilitate the definition of the scenarios that are relevant for answering our research question.

Since our main objective is to provide actionable information by detecting patterns of SSRL using trace data, the next steps are: i) to identify and describe additional collaborative scenarios that illustrate how teachers can benefit from our approach. ii) to identify which types of data sources can help us detect SSRL patterns, as suggested by the learning design of previously detected scenarios; iii) to identify what actionable information we want to generate through SSRL; iv) to explore the use of machine learning techniques (e.g., process mining) to discover SSRL patterns. Then, we have to put them into practice with accessible datasets. It is expected that we will be able to conduct two studies during the second term of this academic year, where we expect to detect SSRL patterns through both different platforms and learning designs.

ACKNOWLEDGEMENTS

This research is partially funded by the European Regional Development Fund and the National Research Agency of the Spanish Ministry of Science, Innovations and Universities under project grants TIN2017-85179-C3-2-R.

REFERENCES

Conole, G. (2013). Designing for learning in an open world. In *Designing for Learning in an Open World*. https://doi.org/10.1007/978-1-4419-8517-0

Er, E., Gómez-Sánchez, E., Dimitriadis, Y., Bote-Lorenzo, M. L., Asensio-Pérez, J. I., & Álvarez-Álvarez, S. (2019). Aligning learning design and learning analytics through instructor involvement: a

MOOC case study. *Interactive Learning Environments*. https://doi.org/10.1080/10494820.2019.1610455

Hadwin, A. F., Järvelä, S., & Miller, M. (2011). *Handbook of self-regulation of learning and performance* (pp. 65–84). Educational psychology handbook series.

Järvelä, S., Gasevic, D., Seppänen, T., Pechenizkiy, M., & Kirschner, P. A. (2020). Bridging learning sciences, machine learning and affective computing for understanding cognition and affect in collaborative learning. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.12917

Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. (2019). What multimodal data can tell us about the students' regulation of their learning process? *Learning and Instruction*. https://doi.org/10.1016/j.learninstruc.2019.04.004

Järvelä, S., Malmberg, J., & Koivuniemi, M. (2016). Recognizing socially shared regulation by using the temporal sequences of online chat and logs in CSCL. *Learning and Instruction*. https://doi.org/10.1016/j.learninstruc.2015.10.006

Jovanovic, J., Dawson, S., Joksimovic, S., & Siemens, G. (2020). Supporting actionable intelligence: reframing the analysis of observed study strategies. *Companion Proceedings 10th International of Conference on Learning Analytics & Knowledge (LAK 2020)*, 161–170. https://doi.org/10.1145/3375462.3375474

Kreijns, K., Kirschner, P., & Jochems, W. (2003). Identifying the Pitfalls for Social Interaction in Computer-Supported Collaborative Learning Environments: A Review of the Research. *Computers in Human Behavior*, *19*, 335–353. https://doi.org/10.1016/S0747-5632(02)00057-2

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing Pedagogical Action: Aligning Learning Analytics With Learning Design. *American Behavioral Scientist*. https://doi.org/10.1177/0002764213479367

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*. https://doi.org/10.1016/j.cedpsych.2017.01.009

Malmberg, J., Järvelä, S., Järvenoja, H., & Panadero, E. (2015). Promoting socially shared regulation of learning in CSCL: Progress of socially shared regulation among high- and low-performing groups. *Computers in Human Behavior*, *52*, 562–572. https://doi.org/10.1016/j.chb.2015.03.082.

Mor, Y., & Craft, B. (2012). Learning design: Reflections upon the current landscape. *Research in Learning Technology*. https://doi.org/10.3402/rlt.v20i0.19196

Peffers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, *24*, 45–77.

Rodríguez-Triana, M. J., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2015). Scripting and monitoring meet each other: Aligning learning analytics and learning design to support teachers in orchestrating CSCL situations. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.12198

Saint, J, Gasevic, D., Matcha, W., Ahmad U., N., & Pardo, A. (2020). Combining analytic methods to unlock sequential and temporal patterns of self-regulated learning. *Companion Proceedings 10th International of Conference on Learning Analytics & Knowledge (LAK 2020)*, 402–411.

Saint, John, Whitelock-Wainwright, A., Gasevic, D., & Pardo, A. (2020). Trace-SRL: A Framework for Analysis of Micro-Level Processes of Self-Regulated Learning from Trace Data. *IEEE Transactions on Learning Technologies*. https://doi.org/10.1109/TLT.2020.3027496

Siadaty, M., Gasevic, D., & Hatala, M. (2016). Trace-Based Microanalytic Measurement of Self-Regulated Learning Processes. *Journal of Learning Analytics*, *3*.

Voogt, J., & Roblin, N. P. (2010). 21st century skills. *Discussienota. Zoetermeer: The Netherlands: Kennisnet*, *23*(03), 2000.

When Gamification meets Learning Analytics

Élise Lavoué
University of Lyon, University Jean Moulin Lyon 3, iaelyon school of Management, CNRS, LIRIS, FRANCE
elise.lavoue@liris.cnrs.fr

Audrey Serna
INSA Lyon, Lyon, France
audrey.serna@insa-lyon.fr

Davinia Hernández-Leo
Universitat Pompeu Fabra, Barcelona, Spain
davinia.hernandez-leo@upf.edu

Katrien Verbert
KU Leuven, Leuven, Belgium
katrien.verbert@cs.kuleuven.be

Vero Vanden Abeele
KU Leuven, Leuven, Belgium
vero.vandenabeele@kuleuven.be

ABSTRACT: This workshop aims to gather researchers from both the Gamification and Learning analytics domains. These two complementary approaches have a common goal: to improve learner motivation and engagement. While the gamification approach tends to integrate motivational mechanisms relevant for learners into learning environments, learning analytics aim at identifying and predicting learner motivation and engagement during a course. Researchers will be invited to present their ongoing projects at the intersection of these two areas, in order to identify the future agenda for the research field.

Keywords: Gamification, Learning analytics, Motivational techniques, Learner engagement

1 ORGANIZERS

Élise Lavoué, Université Jean Moulin Lyon 3, Lyon, France
Audrey Serna, INSA Lyon, Lyon, France
Davinia Hernández-Leo, Universitat Pompeu Fabra, Barcelona, Spain
Katrien Verbert, KU Leuven, Leuven, Belgium
Vero Vanden Abeele, KU Leuven, Leuven, Belgium

Workshop program committee:
Joan Arnedo-Moreno, Universitat Politècnica de Catalunya (UPC), Spain
Manuel Caeiro Rodríguez, University of Vigo, Spain
Baltasar Fernández-Manjón, Universidad Complutense de Madrid, Spain
Sébastien George, Le Mans Université, France
Stuart Hallifax, University Jean Moulin Lyon 3, France
Wilk Oliveira, University of São Paulo, Brazil
Alejandro Ortega-Arranz, Universidad de Valladolid, Spain
Abelardo Pardo, University of South Australia, Australia
Éric Sanchez, Université Fribourg, Switzerland

2 BACKGROUND

On the one hand, gamification is an approach that has been well adopted these past years to enhance learner motivation, performance and engagement with learning environments (Landers & Landers, 2014; Denny et al. 2018). This motivational approach relies on the integration of game design elements in non-game contexts (Deterding et al., 2011), for instance badges, leaderboards, or points, to learning environments (Kapp, 2012). As research in gamification matures, game elements are better understood in terms of the mechanics they support and the user psychological needs they rely on (Jia et al. 2016). Several conceptual and empirical studies on their impact on learners are available, focusing mainly on the impact on learner motivation and performances as final outcomes and less on their engagement as a process (Lavoué et al., 2019).

On the other hand, the Learning Analytics (LA) approach allows to collect, analyze and represent data on learners during a course, based on their interactions with the learning environment, including motivational features (Long & Siemens, 2011). It allows understanding dynamic processes such as learner engaged behaviors, motivation, and emotions during a course (Hernández-Leo et al., 2019; Cukurova et al., 2020).

Thus, these two approaches seem to be complementary and we believe can enrich each other. For instance, the knowledge acquired on learners from LA techniques could be used for adapting and personalizing game elements (Hallifax et al., 2019). Also, designing engaging game elements that reflect student progress, behaviors, and performances could engage learners in reflexive processes. However, little is known yet on how learning analytics can contribute to the research field of gamification, and how gamification can contribute to the learning analytics domain.

3 WORKSHOP OBJECTIVES

We would like to bridge the gap between two research fields and gather a community around the question of how research in Gamification and in Learning Analytics can contribute to each other. This workshop aims to study the possible insights of combining the gamification and the learning analytics approaches in the educational domain, the main issues raised by such a combination and new avenues for future research.

The contributions of the workshop will be published in a joint "LAK Companion Proceedings". About 5-6 contributions are expected for publication. If we do not reach the target number of contributions, we will invite researchers in the field of gamification and learning analytics to present their work in the morning, followed by participatory activities in the afternoon (see details below). We plan to organize a special issue in a journal. The workshop participants will be invited to submit an extended version of their paper.

4 ORGANISATIONAL DETAILS

The expectations of this workshop are described below:

- Type of event: the workshop includes a series of presentations with an open call to present works on gamification and learning analytics with a review process conducted by a committee, including the workshop organizers and other experts. These presentations will be followed by active participatory activities to work together on a research agenda for the future.
- Proposed schedule and duration: full-day.
- Type of participation: 'open' workshop (i.e., any interested delegate may register to attend)
- Workshop activities: The morning will be dedicated to presentations and questions. Participatory activities by groups of 4 or 5 participants will be organized in the afternoon, to work on the main issues and future agenda in the field.
- Planned dissemination activities to recruit attendants: the workshop will be disseminated using the 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.

5 PROPOSED CALL FOR PAPERS

An open Call for Papers will be done for this workshop. Each paper proposal will be reviewed by at least two experts from the Program Committee. The workshop organizers, as experts in both LA and gamification, will do meta-reviews and take the final decisions on acceptance. Contributions on some of the following questions will be welcome:

- What is the current state of the art on gamification and learning analytics in education?
- How can learning analytics techniques be used to evaluate the impact of game elements?
- How can learning analytics be used for adaptive gamification?
- How can game elements support self-regulatory and/or reflexive processes?
- How can learning analytics support teachers in orchestrating game-based learning activities?

- How to design motivational affordances for learning environments? How can LA techniques be used to evaluate these affordances?
- How can gamification and learning analytics techniques complement each other in gamified learning dashboards?

ACKNOWLEDGEMENTS

This workshop is organized as part of the LudiMoodle project financed by the e-FRAN Programme d'investissement d'avenir, operated by the Caisse des Dépots.
D. Hernández-Leo acknowledges the support by ICREA under the ICREA Academia programme and by FEDER and the Spanish Research Agency TIN2017-85179-C3-3-R.

REFERENCES

Denny, P., McDonald, F., Empson, R., Kelly, P., Petersen, A. Empirical Support for a Causal Relationship Between Gamification and Learning Outcomes, in: R. L. Mandryk, M. Hancock, M. Perry, A. L. Cox (Eds.), 24 *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI 2018*, Montreal, QC, Canada, April 21-26, 2018, ACM, 2018, p. 311

Deterding, S. Dixon, D., Khaled, R., and Nacke, L. (2011) From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*. ACM, 9–15.

Hallifax, S., Serna, A., Marty, J. C., & Lavoué, É. (2019). Adaptive Gamification in Education: A Literature Review of Current Trends and Developments. In *European Conference on Technology Enhanced Learning* (pp. 294-307). Springer, Cham.

Hernández-Leo, D., Martinez-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*, 51(1), 139-152.

Jia, Y., Xu, B., Karanam, Y., Voida, S. Personality-targeted gamification: A survey study on personality traits and motivational affordances, in: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, *CHI '16*, ACM, 2016, pp. 2001–2013. Event-place: San Jose, California, USA.

Kapp, K. M. (2012). *The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education*. John Wiley & Sons.

Landers, R. N, Landers, A. K. An empirical test of the theory of gamified learning: The effect of leaderboards on time-on-task and academic performance, *Simulation & Gaming* (2014) 769–785.

Lavoué, E., Monterrat, B., Desmarais, M., George, S. Adaptive gamification for learning environments, *IEEE Transactions on Learning Technologies* 12 (2018) 16–28.

Long P., & Siemens G., (2011), What is Learning Analytics? *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, Alberta, Canada, Feb 2011. ACM, New York, NY, USA. SOLAR (n.d.), Society for Learning Analytics Research. Retrieved from https://www.solaresearch.org/about/what-is-learning-analytics/ on July 2020.

Matcha, W., 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,* 13(2), 226 – 245.

Dynamic gamification adaptation framework based on engagement detection through learning analytics

Stuart Hallifax[1], Audrey Serna[2], Jean-Charles Marty[3], Elise Lavoué[1]
[1]: University of Lyon, University Jean Moulin Lyon 3, iaelyon school of Management,CNRS, LIRIS UMR5205 Lyon France
[2]: INSA de Lyon - CNRS, LIRIS UMR5205 Lyon France
[3]: Université de Savoie Mont Blanc CNRS, LIRIS UMR5205 Lyon France
<name>.<surname>@liris.cnrs.fr

ABSTRACT: Most current adaptive gamification approaches use what is often called a "static" adaptation approach – i.e. game elements are adapted to users once, generally before using the gamified tool, based on a static user profile. On the other hand, dynamic adaptation proposes to adapt game elements based on user behaviour in real time, reacting to variations in user engagement. In this paper, we propose an adaptation framework using an initial static adaptation based on learner profiles, and a dynamic adaptation that uses learning analytics to refine the static adaptation recommendations. The adaptation system is able to observe various learning analytics to estimate learner engagement, to compare to that of other learners, and then to signal to teachers learners that require a change in their gamified environment. We propose a protocol for a future study to test our approach in real conditions, and provide some recommendations for future directions.

Keywords: Gamification, adaptation, interaction log traces, behaviour, engagement, learning analytics

1 INTRODUCTION

Adaptive gamification is the process of tailoring gamification to individual users. Currently, most approaches use "static" adaptation – i.e. they tailor based on individual static user characteristics once. For example in a literature in adaptive gamification for education (Hallifax et al., 2019) found that 13 papers presented some form of static adaptation, and 7 some form of dynamic adaptation. These user characteristics, or profiles, are generally based on information such as preferences for video games (Tondello et al., 2018, 2016), or motivation for education (Hallifax et al., 2020) or even general personality traits (Goldberg, 1992). These approaches only capture the state of the user once, and can fail to take into account differences that occur in users during the usage of the gamified platform. It is particularly true for users' engagement which is a dynamic process that changes overtime during a course (O'Brien & Toms, 2008). This present research is focused on a dynamic adaptation of game elements integrated in a learning environment based on learner engaged behavior.

In this paper, we propose a learning analytics-based approach to estimate and track learner engagement, and offer the basis of a system that can leverage this engagement tracking to signal to teachers when an adaptation should occur. According to this approach, the teachers make the final decision to adapt game elements to learners during the course.

This paper is structured as such: section 2 presents a brief overview of the related literature on dynamic gamification adaptation approaches and methods and tools to analyze engagement. Section 3 presents our general adaptation framework; section 4 presents an application of the general framework to a specific learning context. Finally, in section 5, we conclude, and present future extensions of this work.

2 RELATED WORK

2.1 General dynamic adaptation approaches

There are a few dynamic adaptation approaches in the related gamification literature (Böckle et al., 2017). For example, (Paiva et al., 2016) modelled learner interactions with their learning platform (MeuTutor) to identify patterns in learner behavior to propose personalized missions. They categorized learner actions into four different categories: collaborative (actions related to helping other learners), individual (watching video classes, answering exercises and tests), social (chatting with other learners, sharing class progress), or gamification (achieving various badges and point ranks) interactions. Teachers then used these “interaction profiles” to adapt the learning goals for each individual (i.e. proposing goals that would entice them to interact with otherwise ignored learning content). This resulted in somewhat mitigated effects on learners: on the one hand, this adaptation led to an increase in individual and gamification interactions, but on the other, it failed to have an effect on social interactions. Another example is proposed by (Jagušt et al., 2018) where they present a simple adaptation system. In a math-learning environment, learners fight against a virus by completing mathematics exercises. The system is setup so that the virus “adapts” its speed so that it is always slightly behind the learners at all time. This adaptive setting showed large increases in learner performance, but the authors noted a negative effect on learner stress.

A different interesting approach was proposed by Monterrat et al. (Monterrat et al., 2017, 2015). In their work, they propose to adapt based on various learner profiles (mostly player types) in a static approach. They would then change these initial profiles based on learner behavior – the idea being that would use some kind of framework to link learner behavior and profiles. When the learner’s profile changes by a significant amount, they would reuse the static adaptation rules – i.e. select another game element based on this new learner profile. This approach would be an easy way to implement dynamic adaptation, by enabling the reuse of a breadth of readily available research into links between various learner profiles and relevant game elements. However, to our knowledge there are no framework that link learner behavior (or interaction log traces) to any of the commonly used profiling systems (player profiles, learning styles etc.).

2.2 Explaining and analyzing engagement

Engagement is a complex process. O'Brien & Toms (O’Brien & Toms, 2008) define it as "a quality of user experience characterized by attributes of challenge, positive affect, endurability, aesthetic and sensory appeal, attention, feedback, variety/novelty, interactivity, and perceived user control". It is particularly a dynamic process of engagement and disengagement. To adapt dynamically we need to be able to estimate engagement in real time. Currently, we can identify two major methods of estimating engagement, subjective and objective methods.

Subjective methods generally rely on questionnaires such as User Engagement Scale (UES - (O'Brien et al., 2018)) to determine engagement. However as previously stated, in a dynamic adaptation setting where we need to be able to estimate engagement in real time (or semi-real time), such tools are not quite adequate (i.e. we would not be able to administer multiple questionnaires between usage sessions). Therefore, we can look to more objective methods based on various learning analytics (such as interaction analysis). For example Bouvier et al. (Bouvier et al., 2014) present a study on the engaged behaviors of players in an online sport game. They present a trace model that categorizes user trace actions into four different categories of engagement that are then linked back to the different categories of the Self Determination Theory (Ryan & Deci, 2000). The four categories of engagement identified in this context are: Environmental (linked to *Autonomy* towards the environment), Social (linked to *Relatedness*), Self (*Autonomy* towards the character or role), and Action (linked to *Competence* and *Autonomy* towards the actions). This method was also used for measuring engagement in serious games for education in (LOUP et al., 2016).

We can also look at the methods presented by (Fincham et al., 2019) , who used factor analysis (FA) methods to establish an engagement model based on various simple learning metrics. They list a total of 18 metrics (inspired by the review performed by (Joksimović et al., 2018)). Of these 18, 7 are derived directly from learner trace logs, and are related to behavioural or academic engagement (i.e. days active, question accuracy etc.). The other 11 are related to either cognitive or affective engagement (i.e. learner feeling towards the content, or their knowledge of it). The different engagement categories presented in this paper refer to those distinguished by (Reschly & Christenson, 2012). From these 18 metrics they distinguish a final engagement model based on three engagement factors (using a factor analysis approach). What is interesting is that their factor model does not show a difference between behavioural and academic engagement (something that the authors state is often present in the related engagement literature (Appleton et al., 2008)).

All of the approaches presented here use "*a posteriori*" methods to determine engagement – i.e. after the end of the experiment or usage, so that the researchers can evaluate the impact of the learning system on learner engagement. The challenge for us is to be able to present a method or framework that can estimate learner engagement *during* the usage of a learning environment so that we might be able to adapt game elements according to its evolution (especially if a disengagement is perceived).

3 OUR PROPOSAL – GENERAL ADAPTATION FRAMEWORK

In this paper we propose to develop the adaptation engine architecture presented in Figure 1. This engine is able to select relevant game elements for learners based on both their learner profile, and their behavior (as observed through a set of learning analytics), as presented in blue. This engine therefore employs both a static and dynamic adaptation approach. The static adaptation uses the learner profile and occurs before using the gamified system. It is highlighted in green and described in further detail in section 3.1. The dynamic uses learning analytics and occurs whilst learners are using the gamified system. It is highlighted in yellow, and fully explained in section 3.2.

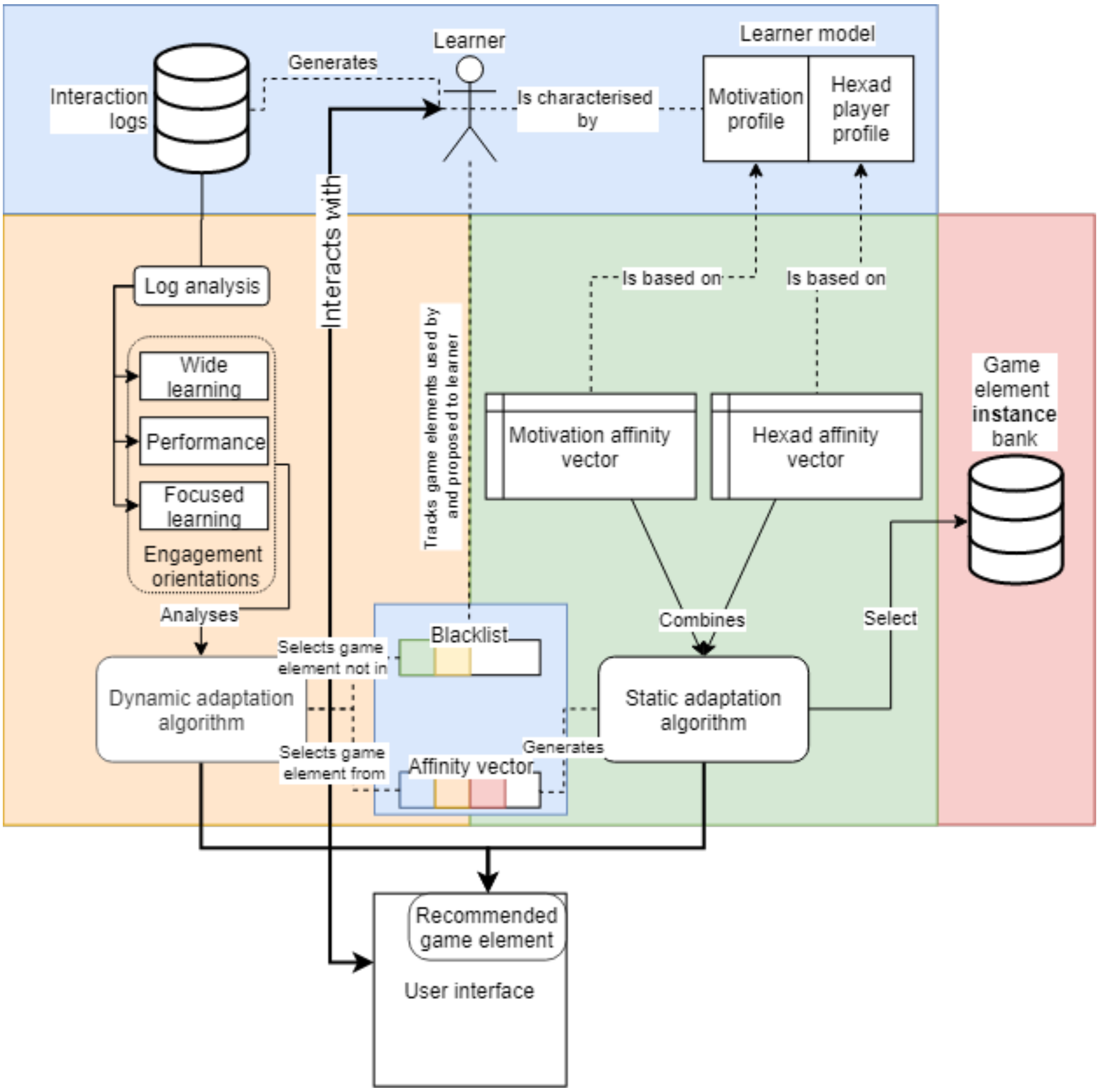


Figure 1 Our adaptation engine architecture. The green section (static adaptation approach) is described in 3.1. The yellow section (dynamic adaptation approach) is described in 3.2. The blue section is the learner information, containing the learner model (combined motivation and player profile), the learner's interaction logs, their affinity vector (an ordered list that represents the most appropriate game elements for the learner generated by the static adaptation algorithm) and their game element blacklist (a list that tracks all the game elements used by and proposed to the learner).

3.1 Initial static adaptation based on learner characteristics

In our study applied to the educational context, (Hallifax et al., 2020) showed that using the Hexad profile (Tondello et al., 2016) as a learner player profile and initial motivation for mathematics (Ryan & Deci, 2000) to statically adapt game element provided significant positive results. We used partial least squares analyses (PLS (Hair Jr et al., 2016), a technique also used in (Orji et al., 2018)) to establish links between the dimensions of the two profiles and effects on learner motivation and behavior. From these analyses, we create an "affinity vector" for each learner (i.e. a list of game elements sorted by descending order of affinity for each learner) which represents how each game element can be expected to positively (or negatively) affect a learner's motivation. We propose to use such a system for the basis of our static adaptation (the method for creating an affinity vector is fully described in our previous work (Hallifax et al., 2020)). Before the learner starts using the gamified learning system, we generate their profile (from questionnaires) and provide them with the most appropriate game element *a priori*.

3.2 Real time dynamic adaptation based on learning analytics

From the initial static adaptation, we can go one step further, and refine these *a priori* predictions and affinities. As previously stated, we can propose a dynamic adaption that takes into account learner behaviour (tracked using learning analytics) to provide an estimation of learner engagement with the learning environment. We propose a dynamic adaptation approach that tracks various types of engagement over the course of the learning experience to suggest an adaptation of the game element when learner engagement decreases. To properly describe how this dynamic adaptation approach should work, we can use the PDA/LPA framework proposed by (Bouzit et al., 2017) – see Figure 2. The PDA/LPA framework describes two adaptation cycles on both the user and the system sides. Users **perceive** an adaptation change, make a **decision** about this change, and **perform** an action (PDA). They then **learn** from this cycle, use this new knowledge to **predict** how the system will react, and **adapt** their behaviour (LPA). This new adapted behaviour flows back into their perceptions and the cycle starts again. On the system side, the system **perceives** learner actions, makes a **decision** based on these perceptions, and performs an adaptation **action** (PDA). The system then **learns** from this adaptation, **predicts** how this will affect the user, and **adapts** its adaptation system (LPA). In our adaptation engine, we propose a simplified version that does not follow all of the PDA/LPA steps, providing a simplified version that allows for future expansion. Notably we do not leverage co-evolution between learner and system, neither does the system predict or adapt to the learner's behaviour. These could be further explored in future expansions of the system. The system works as follows:

1. The learner interacts with the learning platform, generating the interaction log traces (or learning analytics) – **Action**
2. The system analyses these log traces and estimates learner engagement – **Perception**
3. The system decides whether a game element adaptation is required (i.e. if learners show a decrease in engagement) – **Decision**
4. The system proposes a new relevant game element for this learner – **Action**
    a. The learner sees this new game element – **Perception**
    b. They interact with this new game element generating new interaction logs – **Action**
    c. This new game element has an effect on their behaviour and engagement - **Adaptation**
5. On the system side: the proposed game element is blocked for the learner (i.e. it is not reoffered in future) – **Learning**
6. The system then analyses the new log traces generated by the learner and estimates his/her engagement – **Perception** - (restart from step 2)

This adaptation system raises a few questions that need to be answered: (1) How can we control that these adaptations are effective? ((Bouzit et al., 2017) propose that: "controllability is essential to enable the end-user to be actively involved in any adaptation activity") (2) How can we ensure that these adaptations do not occur too often (creating an unstable environment for the learner), and how can we ensure that adaptations do not occur too *rarely* (and therefore not reacting quickly enough to losses of engagement)? We provide first answers to these questions within the framework of the LudiMoodle project described in next section.

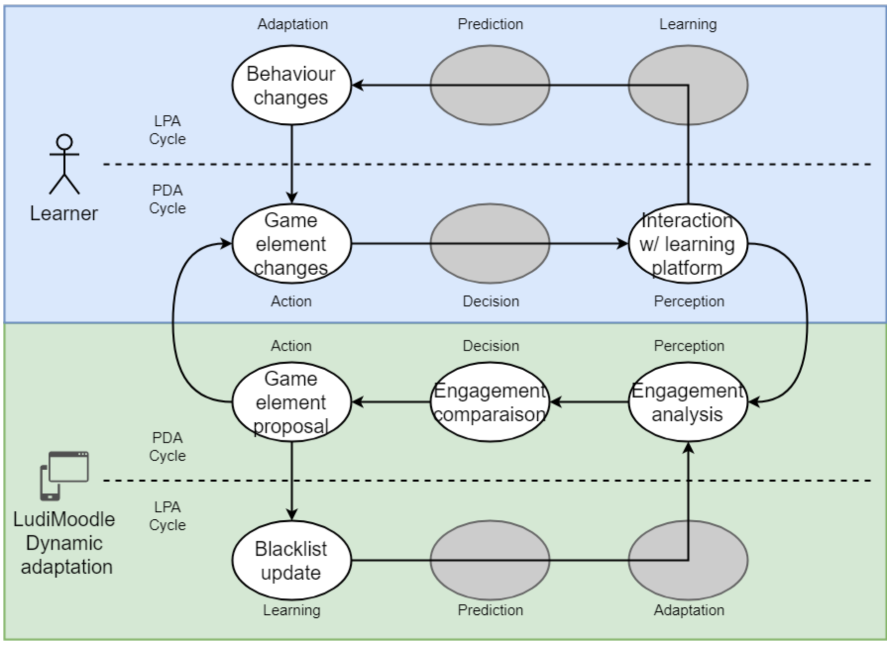


Figure 2 PDA/LPA framework applied to our proposed dynamic adaptation model.

4 APPLICATION TO THE LUDIMOODLE PROJECT

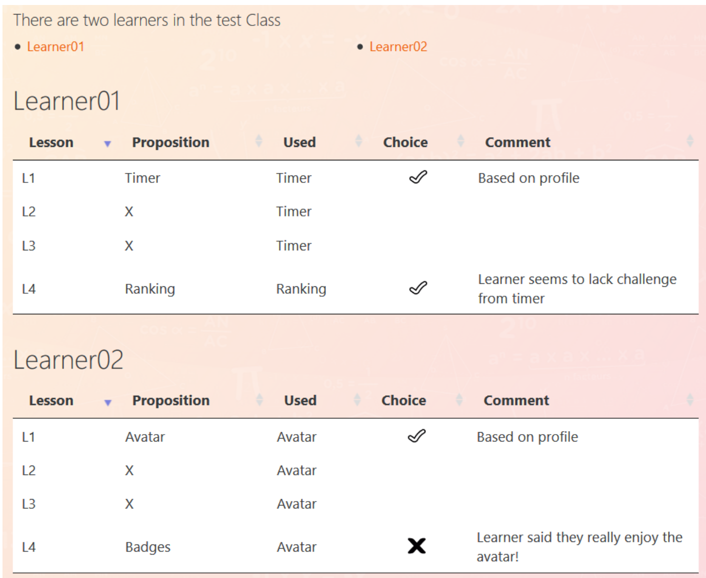
4.1 LudiMoodle project: general approach

In the context of the LudiMoodle project (https://ludimoodle.universite-lyon.fr/), we provided 5 French secondary schools with a gamified learning environment for teaching basic algebra. Data was gathered from 258 learners aged between 13-14 years old over the course of 3 weeks. They used a randomly assigned (i.e. not adapted to their individual profile) game element during 10 lessons. Each of the lessons was composed of between 4 and 10 quizzes of various lengths depending on the complexity of the lesson. The main goal of this project is to compare the effects of (1) statically adapted game elements (2) dynamically adapted game elements and (3) non-adapted (i.e. randomly assigned) game elements. Each of the game elements were chosen to cover one or multiple of the Hexad types as proposed by (Marczewski, 2015). The game elements used in this system are as follows:

- Avatar: Learners have a small goblin character and they can unlock various clothes and accessories for it. Game elements like this one are generally recommended for **Free Spirits**, as these avatars provide them with a personalized representation of themselves.
- Badges: Learners can win various badges based on their lesson progression, individual quiz progression, and longest correct answer streak. Badges are generally shown to be motivating for all users, but should particularly be effective for Players and Achievers, as they represent clear-cut goals for them to achieve with attractive rewards.

- Score: Learners gain points for each correct answer and are shown the maximum point total they can win in each lesson. As this game element gives learners a clear representation of how well they are doing in the course, and rewards them for performing better, it should be attractive to **Players.**
- Timer: Learners are timed for each question. Every time they answer a question correctly, their time is saved. During the next questions of that quiz they are tasked with answering faster, each time that they do, a character runs faster and faster. Here learners are challenged to beat themselves in a race, meaning that **Achievers** should be interested by it.
- Progress Bar: Learners progress in each lesson is represented using a rocket ship that travels towards a planet in space. Each new lesson shows a new planet. This game element should be particularly interesting for the **Achiever** player type as with badges, we also have a clear goal
- Ranking: Learners are placed in a "race" against other learners. Their final position in the race is determined by the number of questions they answered correctly. As this game element allows learners to compare themselves to others, (even if fictional) it should be motivating for **Socializers**

Both the learning content (i.e. quizzes) and game elements were designed in direct collaboration with the teachers who used this tool in their classrooms. This meant that concerning controllability, we proposed that teachers, with their knowledge of: the game elements, the learning content, and the learners; would be the most knowledgeable to understand if the proposed game element is appropriate for a specific learner or not. Between two lessons, teachers are given a table showing the game element each learner is currently using, a suggestion for a new game element that would be more engaging for the learner (if relevant), as well as the previous decisions they made for each learner (if they exist) (see Figure 3).


There are two learners in the test Class

- Learner01
- Learner02

Learner01

Lesson	Proposition	Used	Choice	Comment
L1	Timer	Timer	✓	Based on profile
L2	X	Timer		
L3	X	Timer		
L4	Ranking	Ranking	✓	Learner seems to lack challenge from timer

Learner02

Lesson	Proposition	Used	Choice	Comment
L1	Avatar	Avatar	✓	Based on profile
L2	X	Avatar		
L3	X	Avatar		
L4	Badges	Avatar	✗	Learner said they really enjoy the avatar!

Figure 3 Teacher control interface. In this example, the teacher accepted the proposal for Learner01 (Ranking) and refused the proposal for Learner02 (Badges).

The learning environment automatically stored all learner interaction traces (page visits, question answers, etc.) which we then analysed to determine learner engagement following a similar approach to that proposed by (Bouvier et al., 2014). Our main goal in tracking learner engagement is to be able to detect abnormal decreases, alert the teacher to these decreases and propose a game element adaptation so that we might counter this loss of engagement.

Our analysis proceeded in three steps; first, we reviewed and collated the data available from the LudiMoodle project study using a log trace approach. Second, we ran two factor analyses to create and verify an overarching engagement model that identified the three engagement factors. Finally, we used these factors to track the variation of learner motivation and engagement, and signal to teachers when an adaptation would be necessary. Each step is described in the following sections.

4.2 Determining engagement factors

By analyzing the data that was available from the use of the LudiMoodle platform, we extracted a set of learning analytics that we believed would allow us to follow the evolution of learner engagement and motivation throughout the usage of the system. We then performed two factor analyses (Exploratory FA and Confirmatory FA) to create our final engagement factor model that identifies the following three factors (as calculated in (Hallifax, 2020)):

- **Wide learning engagement** (F1 in figure 4) - This relates to how quickly a learner progressed through the various learning content for a lesson. The more quizzes they passed, the more distinct quizzes they could attempt. Furthermore, the faster they completed each question, the more time they had to attempt other quizzes

- **Performance engagement** (F2) - This directly links to a learner's performance, how well they answered each question and completed each quiz.

- **Focused learning engagement (F3)** - This relates to how much a learner tried to achieve a perfect (100%) score for each quiz, or how much they strived to improve a quiz score.

By calculating these three engagement factors after each lesson, we can estimate how learner engagement varies over time and, more importantly, pinpoint when learners lose engagement. Figure 4 presents an overview of how the different log levels are structured, as well as which operations are used to calculate the different engagement metrics. The final engagement factors along with the metrics that compose them are also displayed. For example, the **performance engagement** factor is computed using the ratio of correct questions (i.e. the percentage of correct questions) and the lesson ratio (i.e. percentage of completed quizzes in a given lesson). The question ratio metric is counted using the complete question operations, which are an aggregation of the *QuizPageView, QuestionGradedWrong/Right, QuizPageView* log traces.

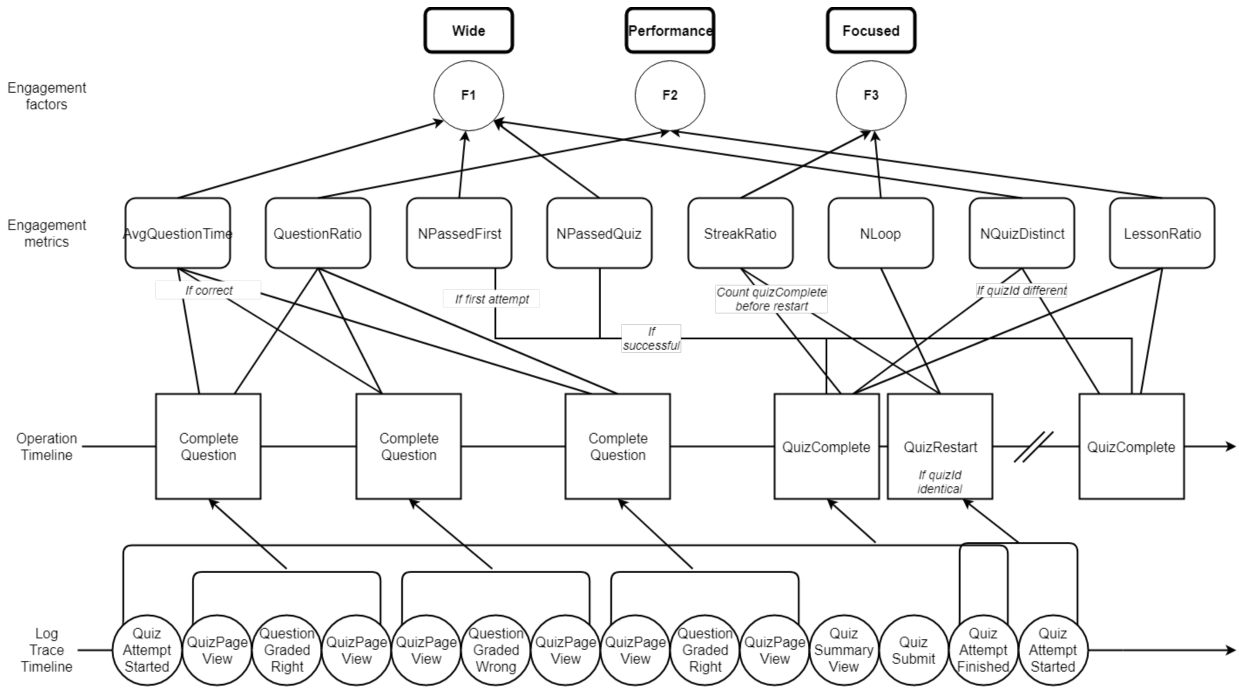


Figure 4. Full log trace analysis model, with the engagement factors (F1, F2, F3) discovered through the EFA/CFA analysis.

4.3 Engagement variations tracking

In our system, we propose to adapt the game element when we detect an abnormal decrease in learner engagement. For each lesson completed, we calculate the intensity of three engagement factors (**Wide learning**, **Performance**, and **Focused Learning**) for each learner, and the variation of these factors with those of the previous lesson. As there is no baseline, or "standard values" for each of these engagement metrics, we decided to compare them to the rest of the learners' class. The idea is that if a single learner displays a decrease in any of the engagement, it is difficult to interpret if it is "normal" or "expected". For example, in the LudiMoodle experiment the later quizzes were harder and more complicated than the earlier ones. This means that it we could expect a slight decrease in performance from all learners, resulting in a decrease in **Performance Engagement**. Therefore, this decrease should not be seen as *exceptional*, and should not trigger a change. This is why we decided to compare learners' variations to those of their classmates. When a change is triggered the system selects the highest game element from the learner's affinity vector that the learner has not yet used. All game elements that the learner has used in the past are blacklisted as to ensure that they are not proposed again in the future.

It is important to note that when a game element is changed, we impose a short cool-off period, where a learner will not be subjected to another adaptation. This is to allow learners to experience their new game element, and get used to it, before a new change could occur. Changing the game element too often could result in confusion in learners. As the teachers in the LudiMoodle project planned to use the platform during ten lessons, we decided to use a cool-off period of three lessons between adaptations as this would results with a maximum of 2-3 game element changes for the least engaged learners. A too short cool-off period could result in an unstable learning environment (frequent changes) and might distract the learner, and too few changes might reduce the system's capacity to react to learner behaviour. Furthermore we use a *blacklist* system to ensure that learners are not offered the same game elements over and over. An example of this approach is presented in Figure 5. During the first three learning sessions, the learner uses the same game element (*blue*). At this time no adaptation is possible, and their blacklist contains one element: *blue* (the one they are currently using). Between the third and fourth learning sessions the system can generate a new game element recommendation: this is the first possible adaptation for this learner. The system computes the variations of the learner's engagement factors (noted ΔW, ΔP, ΔD) between session one and two and between session two and three. These variations are then compared with those of the other learners in their class. In this example, an adaptation is proposed, and the system recommends that the learner use the *purple* game element. The learner's blacklist is therefore updated to contain both *purple* and *blue* meaning that these game elements will not be proposed in the future. In the first timeline, the teacher accepts the adaptation, and the learner is assigned the new game element. They are therefore protected from a further adaptation for the next three sessions. In the second timeline, the teacher refuses, and the learner uses the same game element for session four. The system then uses the variations between S3-S4 and S4-S5 to determine whether an adaptation is required. In this example the system also detects a decrease in engagement, and proposes the *orange* game element. The teacher accepts this new proposal, and the learner is assigned a new game element for the next session.

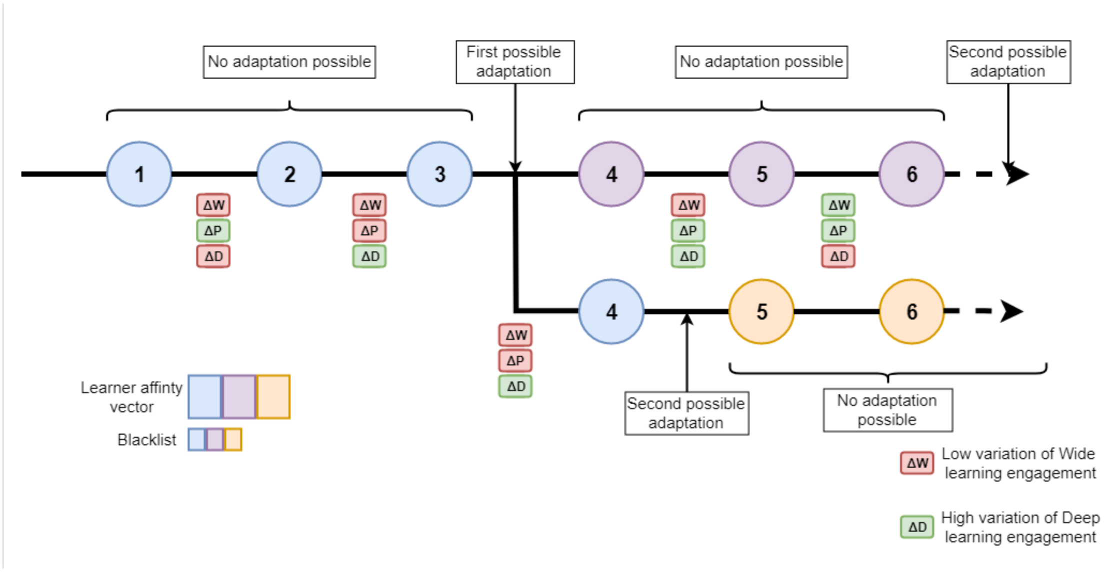


Figure 5. Proposed dynamic adaptation protocol to be used in a real classroom setting

5 CONCLUSION AND FUTURE DIRECTIONS

In this paper we have presented a simple architecture for a game element adaptation engine based on a previously studied static adaptation method (Hallifax et al., 2020) that generates affinity vectors for learners based on their profiles and on a dynamic method that analyses learner behaviour through various learning analytics to estimate learner engagement and refine these static adaptations. Thanks to its controllability property, the system involves teachers directly in the adaptation process, by signalling learners who require an adaptation of the game element they use and suggesting another game element to reengage learners. We believe this approach could be applicable to many contexts, as illustrated within the LudiMoodle project. In its current state, this system is fairly simple and can be expanded in future work to provide recommendations taking into account the context. It would require the system to be able to collect data on the context of the learner activity, and to integrate them in the analysis of their engagement. Currently, this system has not be tested yet in real conditions (the initial field tests had to be interrupted due to pandemic), we plan to conduct an experiment in March-April 2021.

This work opens new avenues for future research. First, the teacher currently only receives a simple notification of which learners need a game element change and a relevant game element proposition for that learner. We could imagine a future version that shows how each learner's engagement evolves over time. This could help teachers reflect on other types of adaptations they could make to the learning system, for example by changing the learning content to better suit the learners. Research in the learning dashboard field could be interesting for example to represent how engagement varies over time (Carrillo et al., 2017).

Second, we could consider a situation where we give control over adaptation to the learners directly. It would allow them to possibly better understand how the adaptation system works and to better predict when adaptation would occur based on their behaviour, and react to it better (i.e. expect the change). This does raise further questions however, for example on the observability and intelligibility of the system. Currently we ask teachers to control adaptation as they have full knowledge of both the game elements and the gamified system, whereas learners will not necessarily be capable to judge which game elements would motivate them more. (Monterrat et al., 2017) showed that learners' appreciation of game elements differs from the observed impact of these game elements on their motivation and engagement. Such a change would also allow us to observe the extent to which the system understands learner preferences.

Third, we could consider improving our learner model by adding more pedagogical related information such as learning styles. (Zaric & Scepanovic, 2018) showed links between various learning styles and different game elements. Our initial static adaptation (that creates the affinity vector) could take into account learning styles. Thus providing our dynamic adaptation through trace analysis a better affinity vector to select game elements from.

Finally, we can enrich our dynamic adaptation approach by including the other steps of the PDA/LPA framework cycles (Bouzit et al., 2017). Currently the dynamic adaptation does not make use of the **Prediction** or **Adaptation** steps. One way we could improve our system is by making it automatically *adapt* to learners. The system could take note of which of the engagement factors decreases the most for each learner (if any) and weight these higher. Currently the system considers each engagement type equally. For example, if a learner loses more **Wide Learning** engagement than the other two, the dynamic adaptation could weight it higher when comparing it to the rest of the class. This means that if the learner lost **Wide Learning** engagement the system would be more likely to propose an adaptation than if they lost **Performance** engagement. This could provide the learner with a more personalised adaptation. However, we would need to be careful as this might result in "uniformising" learners. A solution could be to consult teachers to better understand which factors could be more important to target for adaptation after analysing the interaction data from learners.

6 ACKNOWLEDGEMENTS

This work is a part of the LudiMoodle project financed by the e-FRAN Programme d'investissement d'avenir. We would also like to thank the teachers and students who participated in this study

REFERENCES

Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with school: Critical conceptual and methodological issues of the construct. *Psychology in the Schools*, *45*(5), 369–386.

Böckle, M., Novak, J., & Bick, M. (2017, June). *Towards adaptative gamification: A synthesis of current developments*.

Bouvier, P., Sehaba, K., & Lavoué, É. (2014). A trace-based approach to identifying users' engagement and qualifying their engaged-behaviours in interactive systems: Application to a social game. *User Modeling and User-Adapted Interaction*, *24*(5), 413–451.

Bouzit, S., Calvary, G., Coutaz, J., Chêne, D., Petit, E., & Vanderdonckt, J. (2017). The PDA-LPA design space for user interface adaptation. *2017 11th International Conference on Research Challenges in Information Science (RCIS)*, 353–364. https://doi.org/10.1109/RCIS.2017.7956559

Carrillo, R., Renaud, C., Prié, Y., & Lavoué, É. (2017). Dashboard for monitoring student engagement in mind mapping activities. *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)*, 433–437. https://doi.org/10.1109/ICALT.2017.100

Felder, R. M., & Silverman, L. K. (1988). Learning and teaching styles in engineering education. *Engineering Education*, *78*(7), 674–681.

Fincham, E., Whitelock-Wainwright, A., Kovanović, V., Joksimović, S., van Staalduinen, J.-P., & Gašević, D. (2019). Counting clicks is not enough: Validating a theorized model of engagement in learning analytics. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 501–510.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*(1), 26.

Hair Jr, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage publications.

Hallifax, S. (2020). *Adaptive gamification of digital learning environments.* [Phdthesis, Université Jean Moulin Lyon 3]. https://tel.archives-ouvertes.fr/tel-03125624

Hallifax, S., Lavoué, E., & Serna, A. (2020). To tailor or not to tailor gamification? An analysis of the impact of tailored game elements on learners' behaviours and motivation. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial intelligence in education* (pp. 216–227). Springer International Publishing.

Hallifax, S., Serna, A., Marty, J.-C., & Lavoué, É. (2019). Adaptive Gamification in Education: A Literature Review of Current Trends and Developments. In M. Scheffel, J. Broisin, V. Pammer-Schindler, A. Ioannou, & J. Schneider (Eds.), *Transforming Learning with Meaningful Technologies* (pp. 294–307). Springer International Publishing.

Jagušt, T., Botički, I., & So, H.-J. (2018). Examining competitive, collaborative and adaptive gamification in young learners' math learning. *Computers & Education*, *125*, 444–457.

Joksimović, S., Poquet, O., Kovanović, V., Dowell, N., Mills, C., Gašević, D., Dawson, S., Graesser, A. C., & Brooks, C. (2018). How do we model learning at scale? A systematic review of research on MOOCs. *Review of Educational Research*, *88*(1), 43–86.

LOUP, G., Serna, A., Iksal, S., & George, S. (2016). Immersion and Persistence: Improving Learners' Engagement in Authentic Learning Situations. *11th European Conference on Technology Enhanced Learning, EC-TEL 2016*, 410–415. https://doi.org/10.1007/978-3-319-45153-4_35

Marczewski, A. C. (2015). Even Ninja Monkeys like to play. *CreateSpace Indep. Publish Platform, Charleston*.

Monterrat, B., Lavoué, É., & George, S. (2017). Adaptation of Gaming Features for Motivating Learners. *Simulation & Gaming*, *48*(5), 625–656. https://doi.org/10.1177/1046878117712632

Monterrat, B., Lavoué, É., & George, S. (2015). Toward an Adaptive Gamification System for Learning Environments. In S. Zvacek, M. T. Restivo, J. Uhomoibhi, & M. Helfert (Eds.), *Computer Supported Education* (pp. 115–129). Springer International Publishing.

O'Brien, H. L., Cairns, P., & Hall, M. (2018). A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies*, *112*, 28–39. https://doi.org/10.1016/j.ijhcs.2018.01.004

O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, *59*(6), 938–955. https://doi.org/10.1002/asi.20801

Orji, R., Tondello, G. F., & Nacke, L. E. (2018). Personalizing Persuasive Strategies in Gameful Systems to Gamification User Types. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, *61*, 62.

Paiva, R., Bittencourt, I. I., Tenório, T., Jaques, P., & Isotani, S. (2016). What do students do on-line? Modeling students' interactions to improve their learning experience. *Computers in Human Behavior*, *64*, 769–781.

Reschly, A. L., & Christenson, S. L. (2012). Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. In *Handbook of research on student engagement* (pp. 3–19). Springer.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist*, *55*(1), 68–78.

Tondello, G. F., Valtchanov, D., Reetz, A., Wehbe, R. R., Orji, R., & Nacke, L. E. (2018). Towards a Trait Model of Video Game Preferences. *International Journal of Human–Computer Interaction*, 1–17.

Tondello, G. F., Wehbe, R. R., Diamond, L., Busch, M., Marczewski, A., & Nacke, L. E. (2016). The Gamification User Types Hexad Scale. *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, 229–243. http://dl.acm.org/citation.cfm?id=2968082

Zaric, N., & Scepanovic, S. (2018). Gamification of e-learning based on learning styles–design model and implementation. *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, 723–729.

GamiTool: Towards Actionable Learning Analytics Using Gamification

Alejandro Ortega-Arranz, Alejandra Martínez-Monés,
Juan I. Asensio-Pérez, Miguel L. Bote-Lorenzo
GSIC-EMIC Research Group, Universidad de Valladolid, Spain
alex@gsic.uva.es

ABSTRACT: Learning Analytics enable a better understanding of teaching and learning processes by identifying and monitoring indicators based on students' activity. These same indicators can also be used by reward-based gamification strategies as conditions that students should satisfy to earn rewards, with the purpose of increasing their engagement with the learning contents and activities. Hence, gamification systems must enable the digital representation and interpretation of indicators based on students' activity, similarly as learning analytics tools do. This position paper introduces GamiTool, a gamification system to support the design and the computer-interpretable representation of a wide variety of learning analytics indicators that can be configured by practitioners as gamification conditions. Additionally, the paper discusses five potential lines of work regarding joint research with GamiTool and LA.

Keywords: Learning analytics, gamification, reward-based strategies, indicators, practitioners.

1 INTRODUCTION

Nowadays, playing games is one of the most popular forms of worldwide entertainment (OppenheimerFunds, 2018). Games' entertainment has its roots on the human feelings (*e.g.*, curiosity, excitement, competition) that game designs can generate as part of the interaction with the players. Gamification aims at identifying and implementing in non-game contexts, game design elements (*e.g.*, leaderboards, maps) and techniques (*e.g.*, onboarding, increasing difficulty) able to motivate the users, hold their interest and/or challenge them to solve problems (de Sousa Borges et al., 2014; Simões et al., 2013). One of the contexts where gamification has attracted lot of attention during the last years is online education (Antonaci et al., 2019).

Previous literature reviews on gamification showed that rewards (*e.g.*, experience points) are the game design elements most implemented in online educational environments (Antonaci et al., 2019; Dicheva et al., 2015). These rewards are visual elements (*e.g.*, ribbon) or privileges (*e.g.*, unlock content) that are issued when conditions defined beforehand are fulfilled (Ortega-Arranz et al., 2019). For instance, students can get a deadline extension for a task (*privilege*) when submitting three optional course tasks before a configured deadline (*condition*). Literature has also reported the potential of these reward-based gamification strategies in online environments regarding the improvement of learners' motivation, engagement and learning outcomes, among other benefits (Domínguez et al., 2013; Ibañez et al., 2014; Anderson et al., 2014).

However, the inclusion of reward-based strategies implies a set of orchestration tasks (Prieto et al., 2014) which must be carried out by practitioners (*i.e.*, teachers, instructors, instructional designers).

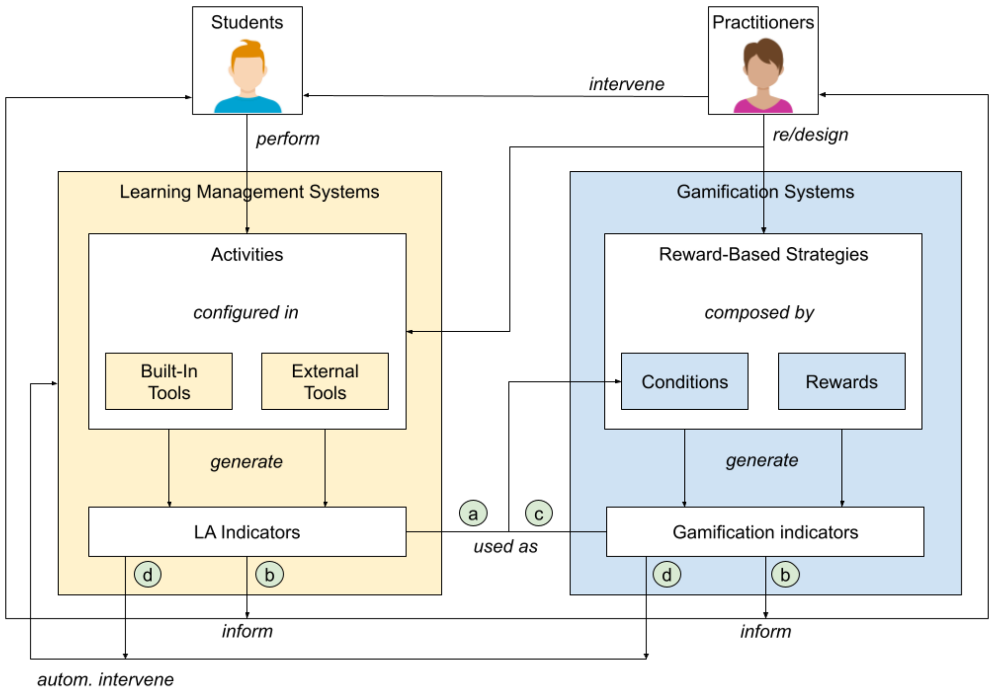


Figure 1: Association between LA and reward-based gamification in online learning environments.

For instance, the translation of practitioners' gamification purposes into reward-based strategies, the configuration of such strategies in the learning platforms, or their management during course run-time. Multiple gamification tools and systems have been created to support practitioners in performing such tasks and alleviate the associated workload such as OneUp (Dicheva et al., 2019), MEdit4CEP-Gam (Calderón et al., 2018) or Badgr[1]. In this situation, gamification systems are expected to provide automatic capabilities for orchestration tasks (*e.g.*, checking gamification conditions) and therefore, capabilities to computationally understand the different components shaping the gamification designs (e.g., course activities, conditions, rewards).

From the Learning Analytics (LA) perspective, reward-based strategies can be conceived as a form of making LA indicators actionable (Dichev et al., 2018), and gamification systems, as LA-design tools able to script, interpret and automate these actionable indicators. These gamification systems enable practitioners decide which indicators and thresholds will be used as conditions (e.*g.*, complete more than 3 peer reviews), and which actions will be taken once the conditions are satisfied (*e.g.*, unlock new learning contents). Usually, these conditions are based on students' behavioral data obtained from the system logs (see *(a)* in Figure 1) and represent students' actions considered beneficial by practitioners for the purposes for which gamification is used. For instance, counting students' posts to foster interaction in discussion forums (Anderson et al., 2014) or engage in practice by repeating exercises (Dicheva et al., 2019). These indicators can range from very simple information, close to raw data (*e.g.*, number of posts in forums), to more sophisticated ones, based on natural language processing or advanced analytic techniques (*e.g.*, thoughtful users) (Ruipérez-Valiente et al., 2015).

---

[1] Concentric Sky. Badgr: https://info.badgr.com/, last access: March, 2021.

This paper presents GamiTool (Section 2), a gamification system developed by the authors to support practitioners in the orchestration of reward-based strategies. This system supports: (a) the creation of flexible computer-interpretable gamification designs; (b) the automatic implementation and enactment of gamification designs according to the values and thresholds of LA indicators configured by practitioners; and (c) the real-time visualization of gamification indicators. Furthermore, Section 3 introduces some future lines of research considering GamiTool and the last advances carried out in the LA field. Finally, Section 4 outlines some conclusions obtained from this work.

2 GAMITOOL: SUPPORTING THE ORCHESTRATION OF REWARD-BASED GAMIFICATION STRATEGIES

GamiTool[2] is a gamification system implementing an adapter-based architecture to support the orchestration of reward-based strategies in multiple learning management systems (Ortega-Arranz et al., *under review*). GamiTool also incorporates a gamification-specific data model to support the flexible design of reward-based strategies (*i.e.*, conditions and rewards) as conceived by practitioners (see Figure 2). This flexibility in the design of gamification strategies aims at enabling practitioners to select a wide variety of LA indicators of their own choice to adapt their gamification conditions to their learning designs.

More specifically, GamiTool enables the computer-interpretable representation of gamification designs involving 10 different types of learning resources (*e.g.*, assignments, quizzes, content pages), and 30 different actions (*e.g.*, log in, submit, post) which can be further specified into multiple fine-grain rules (*e.g.*, before a specific date, several times). For instance, a gamification condition can be configured as *students must submit (action) the questionnaire located at Module 1 (resource) before the configured deadline (rule) and score higher than 90% (rule) in the first attempt (rule)*. This gamification condition shows the use of multiple behavioral indicators that can potentially inform about those students that are more engaged with the contents (first attempt, high score) at the envisioned course pace (condition deadline).

Apart from fine-grained conditions, GamiTool also supports the gamification of actions that must be satisfied by a specific percentage of group members (*e.g.*, at least 50% of group peers must contribute to the collaborative group glossary). Therefore, group activities can be also gamified to foster the individual accountability of group peers to achieve the reward. Additionally, since many learning situations are configured in distributed learning environments (Prieto et al., 2014), GamiTool was designed to interact with both LMS built-in and external tools, being able to retrieve behavioral indicators from both learning management systems (*e.g.*, Moodle, Canvas) and external tools (*e.g.*, Google Spreadsheets) in a same gamification design. In summary, all these features for the design of computer-interpretable gamified designs provide practitioners with a high flexibility when configuring LA indicators as gamification conditions (*e.g.,* conditions based on student individual actions, based on group actions, based on peer approval, based on previous earned rewards).

[2] Further information about GamiTool can be found at https://www.gsic.uva.es/gamitool/, last access: March 2021.

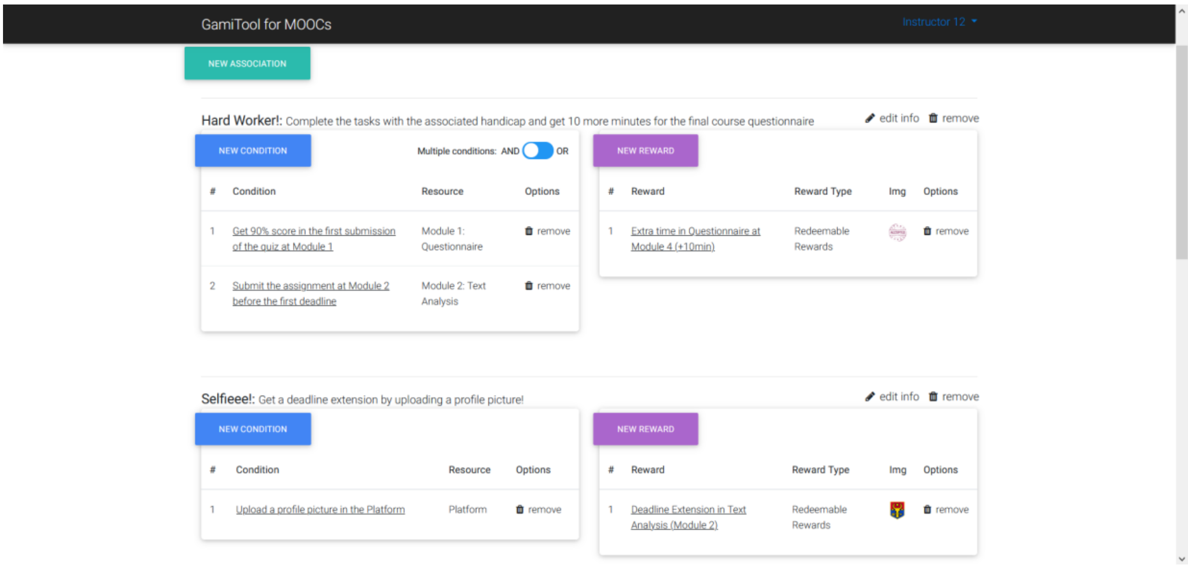


Figure 2: Screenshot of the GamiTool association page.

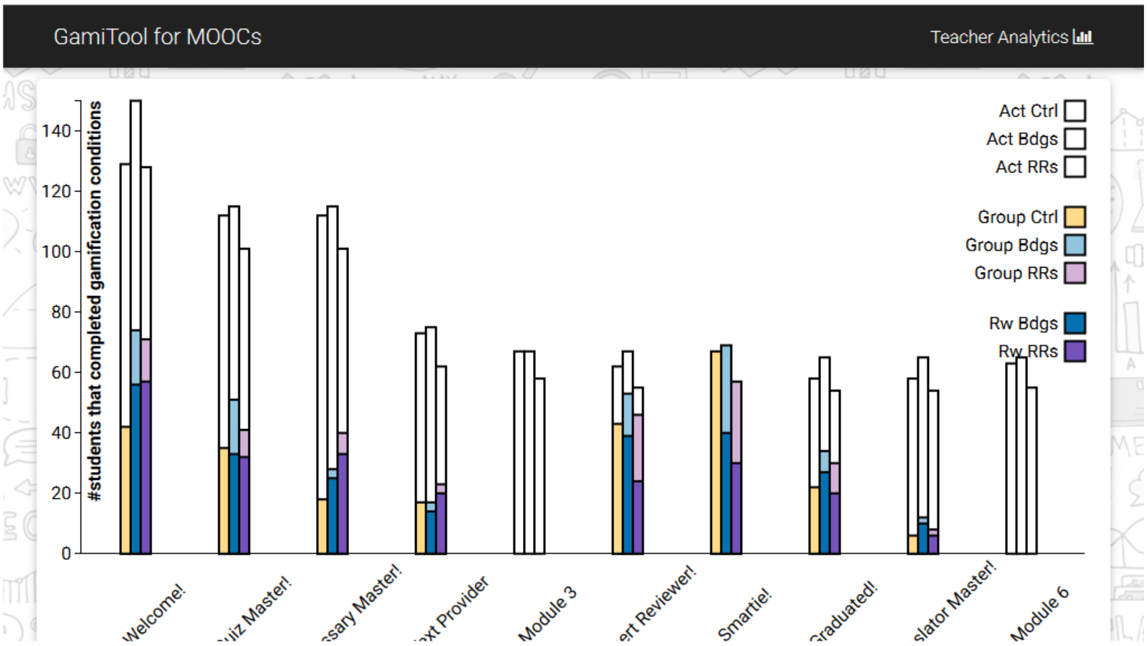


Figure 3: Screenshot of the GamiTool analytics page.

In addition, as represented in Figure 1, the interaction of students with the gamified activities, and therefore, with the configured reward-based strategies, generates multiple gamification indicators such as the number of earned rewards, the student positions in the leaderboards, the number of students that earned a concrete reward, etc. (Heilbrunn et al., 2017; Dichev et al., 2018). The gamification analytics strongly depend on the decisions taken during the design of gamification (*e.g.*, those designs including leaderboards can track the current position and the progress of the student in the leaderboard). These gamification indicators (see Figure 3) are potentially useful for monitoring and re-designing reward-based strategies, *e.g.*, a gamification condition is so difficult that only few students earned it (see *(b)* in Figure 1). Like the LA indicators, gamification indicators can be also used for the definition of new gamification conditions (*e.g.*, a new reward will be issued when student rank first in the leaderboard or when students earn 100 experience points) and for automatic interventions in the learning design of the course (see *(c)* and *(d)* in Figure 1, respectively). This type of conditions based on gamification analytics are also supported by GamiTool.

A prototype of GamiTool has been developed supporting part of its functionality in two learning management systems (*i.e.*, Moodle, Canvas) and one external tool (*i.e.*, Google Spreadsheets). This prototype was used in a real MOOC hosted at Canvas Network (Ortega-Arranz et al., 2019), and was evaluated by 19 practitioners and gamification designers from multiple institutions (Ortega-Arranz et al., *under review*). In this last evaluation, among other features, practitioners were requested to create a gamification design with reward-based strategies over a given MOOC. Therefore, we could understand the extent to which GamiTool supports the representation of computer-interpretable gamification designs as configured by real practitioners and gamification designers (Ortega-Arranz et al., *under review*). Results showed that from all the configured gamification conditions (N=71), GamiTool can directly represent most of them (59.15%), including those involving behavioral indicators such as submitting quizzes, getting likes in discussion forums, or completing rubrics for peer review. On the other hand, results also revealed that a considerable percentage of conditions (39.44%) could be quantitatively supported by GamiTool but, ideally, they would require content analysis (*e.g.*, submit three economical terms to the course glossary) or more complex indicators (*e.g.*, active participation in the group task). Support to these two types of conditions has been outlined in the GamiTool roadmap agenda.

Another important feature of the evaluation was to understand the extent to which GamiTool helped practitioners to create gamification designs. In this situation, the GamiTool catalog of supported behavioral indicators that can be tracked for the different course activities was, in general, perceived as useful for the definition of new conditions, and showed a moderate positive correlation ($r(18)= -0.470, p= 0.049$) with the previous gamification experience of the users (Ortega-Arranz et al., *under review*). The evaluation also showed an "excellent" usability of the system (Brooke, 1996) according to SUS score ratings (*mean(19)= 84.61*), and a low perceived workload (Hart, 2006) for designing and implementing gamification designs according to the *Raw TLX* score ratings (*mean(19)= 31.57*).

3 EXPLORING SINERGIES BETWEEN GAMITOOL AND LA

LA (Siemens & Long, 2011) and gamification (Deterding et al., 2011) have been proposed to address similar educational drawbacks (*e.g.*, improving student learning outcomes, fostering self-regulated learning) by using similar indicators. However, both research areas created their own agenda separately. As a case of this separate trends, the research work leading the development of GamiTool was mainly inspired in the gamification-related literature and did not consider its connections with the learning analytics field. Consequently, we have made ourselves two questions: *How can GamiTool benefit from the LA field?* and the other way around, *What could be the contribution of GamiTool to the LA field?* These questions have led to five potential lines of work regarding the relation between GamiTool and LA.

Extending the Use of Gamification Analytics: As mentioned before, the student interaction with the reward-based strategies generates new variables that can be tracked by GamiTool (*e.g.*, time span for claiming rewards, number of earned rewards, position in the leaderboard, students completing a configured condition). These gamification analytics can potentially complement the traditional learning analytics provided by learning management systems for multiple purposes (*e.g.*, students that are most and least active, engagement level of tasks). Another envisioned purpose is the use of these variables as input variables of LA frameworks to model and predict student behaviors more accurately (Ranjeeth et al., 2020). Furthermore, modelling student behavior considering both learning and gamification analytics could help define personalized learner and player profiles more accurately, thus supporting the effectiveness of tailored gamification (Hallifax et al., 2020). As a future work, we plan to understand the relationship between these gamification variables and the student behaviors (*e.g.*, task submission, dropout) in our previous studies about reward-based strategies in MOOCs, and to propose their integration within existing LA models.

Complex Indicators as Conditions: GamiTool supports the configuration of gamification conditions based on indicators of student activity (*e.g.*, post a comment in a discussion forum) which can be further combined with other indicators to form more complex conditions (*e.g.*, upload a profile picture, post a comment in the discussion forum, and receive at least 5 likes in such comment from course peers). Currently, GamiTool supports a set of basic indicators which were shown to be the ones most frequently envisioned by practitioners in our evaluation. However, a considerable number of conditions proposed by the participants of the evaluation studies also involved content analysis (*e.g.*, "submit three economical terms to the course glossary") and more complex indicators (*e.g.*, "active participation in group work"). Therefore, considering the existence of LA tools able to obtain complex LA indicators (*e.g.*, Khalil & Belokrys, 2020), it seems interesting to study the connection of GamiTool with such LA tools and to assess their usefulness for practitioners.

LA Frameworks for Gamification Design: One of the outcomes of the evaluation performed with practitioners and gamification designers was the usefulness of GamiTool to inspire the design of new forms of gamification, thanks to the fact it displays all the supported indicators for every resource type of the learning design. In this same evaluation, practitioners also raised the potential usefulness of being supported in the design regarding the most suitable number of rewards (and conditions), and/or the most important resources to be gamified according to the purposes for which gamification is being implemented (*e.g.*, increase learning outcomes, foster participation). Given this context, there exist LA frameworks providing useful insights about how LA indicators relate to different learning purposes (*e.g.*, Gašević, et al., 2017). Therefore, it seems interesting to explore to what extent these frameworks can be transferred to the context of gamification, and if such frameworks also produce a meaningful support in the design of gamification conditions. As a future work, the conceptual elements of such frameworks could be combined with GamiTool-DM to explore their applicability and usefulness within GamiTool.

Gamification Across-Spaces: GamiTool currently supports the gamification of learning management systems and external tools, thus supporting the creation of gamification designs that involve multiple virtual learning environments and tools (*e.g.*, Moodle, Canvas, Google Spreadsheets). However, learning does not only occur in the digital space but also in the physical one. Gamification experiences in the physical space normally require the manual monitoring and management of student behavior.

Nevertheless, the rapid growth of sensors in the last decade (*e.g.*, smartphone sensors) enable their use to automatize data collection from the physical space. During the last years, there has been much research regarding this aspect in the LA field under the umbrella of the so-called "CrossMMLA" (Giannakos et al., 2020). Therefore, future refinement of GamiTool should leverage the current state of CrossMMLA to support reward-based gamification in across-spaces learning situations (*e.g.*, GamiTool-DM extension to support the gamification of actions performed in the physical space).

Ethics and Data Privacy: The collection and measurement of student information involves some ethical and private issues that are more aggravated when this information is not only used by learning management systems but also by third-party gamification systems. The ethical and privacy implications of tracking students' actions are been extensively addressed by the LA research field, including the development of educational systems (*e.g.*, Hoel et al., 2017). However, this does not happen in gamification. Further work is needed to understand whether the policies proposed in the LA area are transferable to the concrete case of gamification and gamification systems, including GamiTool.

4 CONCLUSIONS

This paper reflects about the (bi-directional) relationship between LA and GamiTool, a gamification system initially conceived under the frame of gamification-based research. GamiTool enables the definition of gamification conditions at design time, based on indicators of student activity, and which can represent goals expected to be achieved as defined by practitioners. The definition of indicators considering the Learning Design relates to the current efforts from the LA field to script actionable LA at design time (*i.e.*, rewards and privileges are given to the student according to conditions considered beneficial by practitioners).

On the one hand, latest LA advances can provide to GamiTool (and to other gamification systems) with design frameworks, tools, and policies, supporting the design of effective gamification strategies, the incorporation of more complex gamification conditions and the secure treatment of student personal data. On the other hand, GamiTool can provide to the LA field new gamification variables potentially defining the behavior of students with learning resources and activities. This information could be used together with traditional learning analytics to better understand and predict the learning processes happening in online courses and generate automatic interventions that can alleviate the associated orchestration workload to practitioners. All these lines point out to potential lines of future work combining the research fields of LA and gamification.

ACKNOWLEDGEMENTS

This research has been partially funded by the Spanish State Research Agency (AEI) together with the European Regional Development Fund, under project grant TIN2017-85179-C3-2-R; the Regional Government of Castilla y León together with the European Regional Development Fund, under project grant VA257P18; and, the European Commission, under project grant 588438-EPP-1-2017-1-EL-

EPPKA2-KA. The authors thank the rest of the GSIC-EMIC team for their valuable ideas and support received conducting this research.

REFERENCES

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014). Engaging with massive online courses. In *Proceedings of the 23rd international conference on World wide web* (pp. 687-698).

Antonaci, A., Klemke, R., & Specht, M. (2019). The effects of gamification in online learning environments: A systematic literature review. *Informatics* (Vol. 6, No. 3, p. 32). Multidisciplinary Digital Publishing Institute.

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. (pp. 189-194). London: Taylor & Francis.

Calderón, A., Boubeta-Puig, J., & Ruiz, M. (2018). MEdit4CEP-Gam: A model-driven approach for user-friendly gamification design, monitoring and code generation in CEP-based systems. Information and Software Technology, 95, 238-264.

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: defining" gamification". In *Proceedings of the 15th international academic MindTrek Conference: Envisioning future media environments* (pp. 9-15).

Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: A systematic mapping study. *Journal of Educational Technology & Society*, 18(3), 75-88.

Dicheva, D., Irwin, K., & Dichev, C. (2019). OneUp: Engaging students in a gamified data structures course. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (pp. 386-392).

Dichev, C., Dicheva, D., & Irwin, K. (2018). Gamification Driven Learning Analytics. In *Proceedings of the 9th International Conference on E-Learning* (pp. 70-76).

Domínguez, A., Saenz-de-Navarrete, J., De-Marcos, L., Fernández-Sanz, L., Pagés, C., & Martínez-Herráiz, J. J. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers & education*, 63, 380-392.

Gašević, D., Kovanović, V., & Joksimović, S. (2017). Piecing the learning analytics puzzle: A consolidated model of a field of research and practice. *Learning: Research and Practice*, 3(1), 63-78.

Giannakos, M., Spikol, D., Molenaar, I., Di Mitri, D., Sharma, K., Ochoa, X., & Hammad, R. (2020). Preface: CrossMMLA in practice: Collecting, annotating and analyzing multimodal data across spaces. In *CEUR Workshop Proceedings* (Vol. 2610).

Hallifax, S., Lavoué, E., & Serna, A. (2020). To tailor or not to tailor gamification? An analysis of the impact of tailored game elements on learners' behaviours and motivation. In Proceedings of the 21th International Conference on Artificial Intelligence in Education (pp. 216-227).

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 50, No. 9, pp. 904-908).

Heilbrunn, B., Herzig, P., & Schill, A. (2017). Gamification analytics—methods and tools for monitoring and adapting gamification designs. In *Gamification* (pp. 31-47). Springer, Cham.

Hoel, T., Griffiths, D., & Chen, W. (2017). The influence of data protection and privacy frameworks on the design of learning analytics systems. In *Proceedings of the seventh international learning analytics & knowledge conference* (pp. 243-252).

Ibañez, M. B., Di-Serio, A., & Delgado-Kloos, C. (2014). Gamification for engaging computer science students in learning activities: A case study. *IEEE Transactions on learning technologies*, 7(3), 291-301.

Khalil, M., & Belokrys, G. (2020). OXALIC: an Open edX Advanced Learning Analytics Tool. In *Proceedings of the 2020 IEEE Learning with MOOCs* (pp. 185-190). IEEE.

OppenheimerFunds (2018). Investing in the Soaring Popularity of Gaming. *Retrieved from: https://www.reuters.com/sponsored/article/popularity-of-gaming*, last access: February 2021.

Ortega-Arranz, A., Asensio-Pérez, J. I., Martínez-Monés, A., Bote-Lorenzo, M. L., Ortega-Arranz, H., & Kalz, M. (under review). GamiTool: Supporting Practitioners in the Use of Gamification Strategies for MOOCs. *IEEE Transactions on Learning Technologies*.

Ortega-Arranz, A., Bote-Lorenzo, M. L., Asensio-Pérez, J. I., Martínez-Monés, A., Gómez-Sánchez, E., & Dimitriadis, Y. (2019). To reward and beyond: Analyzing the effect of reward-based strategies in a MOOC. *Computers & Education*, 142, 103639.

Prieto, L. P., Asensio-Pérez, J. I., Muñoz-Cristóbal, J. A., Jorrín-Abellán, I. M., Dimitriadis, Y., & Gómez-Sánchez, E. (2014). Supporting orchestration of CSCL scenarios in web-based Distributed Learning Environments. *Computers & education*, 73, 9-25.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, 46(5), 30.

Simões, J., Redondo, R. D., & Vilas, A. F. (2013). A social gamification framework for a K-6 learning platform. *Computers in Human Behavior*, 29(2), 345-353.

de Sousa Borges, S., Durelli, V. H., Reis, H. M., & Isotani, S. (2014). A systematic mapping on gamification applied to education. In *Proceedings of the 29th annual ACM symposium on applied computing* (pp. 216-222).

Ranjeeth, S., Latchoumi, T. P., & Paul, P. V. (2020). A survey on predictive models of learning analytics. *Procedia Computer Science*, 167, 37-46.

Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., Leony, D., & Delgado-Kloos, C. (2015). ALAS-KA: A learning analytics extension for better understanding the learning process in the Khan Academy platform. *Computers in Human Behavior*, 47, 139-148.

An analysis of the Game Mechanics and Learning Analytics behind Pyramid collaboration scripts

René Lobo-Quintero
Department of Information and Communication Technologies, Universitat Pompeu Fabra
renealejandro.lobo@upf.edu

Davinia Hernández-Leo
Department of Information and Communication Technologies, Universitat Pompeu Fabra
davinia.hernandez-leo@upf.edu

ABSTRACT: Collaborative learning flow patterns (CLFPs) formulate good practices for scripts orchestrating activity sequences and collaboration mechanisms that can elicit fruitful social interactions. Despite their benefits, it is worth exploring how their implementation can be improved. Learning technology research suggests that the use of gamification strategies accompanied with learning analytics offers potential to reinforce the productive participation and collaboration between participants, and at the same time to help teachers to make regulation decisions and measure the impact of the activities. In this paper, we analyze three case studies where the Pyramid CLFP is used. The analysis shows that implementations of this pattern already incorporate different uses of game mechanics and learning analytics. The paper also discusses the approaches in use and how complementary mechanisms could be considered for the further improvement of future designs.

Keywords: Collaborative Learning, CLFP, Gamification, Learning Analytics, Pyramid, Scripts

1 INTRODUCTION

Collaborative Learning Flow Patterns (CLFPs) are topic-independent structures of potentially effective scripted sequences of learning activities that can be adapted to multiple educational scenarios (Hernández-Leo et al., 2006). These patterns can help teachers design and incorporate scripted collaborative learning scenarios into their teaching practice. Certainly, the application of CL poses certain challenges and drawbacks (Radkowitsch et al., 2020), such as (1) students usually divide the tasks, working individually without collaborating; (2) these activities require extra time for both teachers and students; and (3) can emerge eventual interaction and communication problems among students, and CLFPs largely contributes to overcome these challenges (Hernández-Leo et al., 2005), CLFPs achieve so through scripting mechanisms (in the formulation of groups, roles, resource sharing, and activity sequencing) that promote positive interdependence, individual accountability and knowledge sharing (Hernández-Leo et al., 2005, Johnson et al., 2016, Nah et al., 2014). Yet, given the relevance of their aims, additional strategies to reinforce CLFP effects on collaboration and learning are worth studying. In this paper, we explore how game mechanics and learning analytics can enhance the implementation of the scripting mechanisms in CLFPs.

On the one hand, previous research in game-based learning suggests that game mechanics can reinforce a pedagogical strategy able to support several underpinning CL mechanics present in CLFPs, facilitating higher levels of motivation, participation and more enjoyable learning experiences

(Johnson et al., 2016, Nah et al., 2014). Indeed, several studies have addressed the use of game-based learning in collaborative learning activities and environments. For instance, Darejeh & Salim, (2016) propose an ontology to represent gamification strategies in collaborative learning scenarios. Also, Johnson et al., (2016) describe a case study in which an online gamified discussion forum increased student collaboration and reduced response times. On the other hand, the implementation of game mechanics can benefit from the use of learning analytics (Freire et al., 2016). It is expected that the inclusion of game mechanics in the scripts based on the Pyramid CLFP generates key performance indicators so learning analytics can be applied to track and inform teachers about the learning process.

In order to tackle the objective of understanding how game mechanics and learning analytics can enhance the implementation of CLFPs, this paper focuses the study on a specific CLFP that incorporates key scripting mechanisms (dynamic changes in group formation - of increasing size - across a sequence of collaborative learning activities), the Pyramid CLFP (Hernández-Leo et al., 2006). The study is approached through an analysis, guided by the LM-GM framework (Arnab, et al., 2015), of three cases implementing the Pyramid CLFP as reported in the literature.

Thus, the research questions guiding this work are: (RQ1) What game mechanics are present in learning scenarios already implementing Pyramid CLFP scripts? (RQ2) How is learning analytics present in the implementation of Pyramid CLFP scripts? (RQ3) What can be proposed to extend game mechanics and the use of learning analytics in the implementation of Pyramid CLFP scripts?

The next sessions describe the background of the Pyramid CLFP, the LM-GM framework and the notions of gamification and learning analytics. Section 3 summarizes the three implementations of the Pyramid CLFP selected for the analysis. Then, section 4 presents the analysis of the game mechanics present in the selected cases and section 5 focuses on the analysis of the Learning Analytics used. Some proposals to extend the implementation of Pyramid CLFP using game mechanics are discussed in section 6.

2 BACKGROUND

2.1 Collaborative Learning Flow Patterns

Collaborative Learning Flow Patterns (CLFPs) represent broadly accepted techniques that are repetitively used by collaborative learning practitioners (e.g., teachers) when structuring the flow of types of learning activities involved in collaborative learning scenarios (Hernández-Leo et al., 2005). CLFP pre-structure collaboration in such a way that productive interactions are promoted, so that the potential effectiveness of the educational situation is enhanced (Jermann et al., 2004), fostering individual participation, accountability and balanced positive interdependence. Examples of CLFPs include Jigsaw, TPS (Think-Pair-Share), Simulation, TAPPS (Thinking Aloud Pair Problem Solving) and Brainstorming (Hernández-Leo et al., 2006). This paper is focused on analyzing Computer Supported Collaborative Learning (CSCL) studies that used the Pyramid pattern, which is a CLFP with complex scripting structures that cover key scripting mechanisms related to changes in group formation changing (in terms of members and group size) along a flow of several learning activities.

The “Pyramid” pattern is used for complex problems, usually without a specific solution whose process of resolution can benefit from a gradual discussion and consensus among all participants (Hernández-Leo et al., 2005). A Pyramid flow is usually initiated with individual students solving a global task. Then, in a second phase of the Pyramid, such individual solutions are discussed in small groups and agreed upon a common proposal. These small groups then form larger-groups iteratively and large group discussions will continue until a consensus is reached at the global level. Pyramid flows foster individual participation, accountability and balanced positive interdependence (Hernández-Leo et al., 2006). Furthermore, the Pyramid pattern promotes conversations in incrementally sized groups, clear expectations of reaching consensus and positive reinforcement mechanisms leading to desired positive behaviors in the learning process (Fluke & Peterson, 2013; Manathunga & Hernández-Leo, 2018).

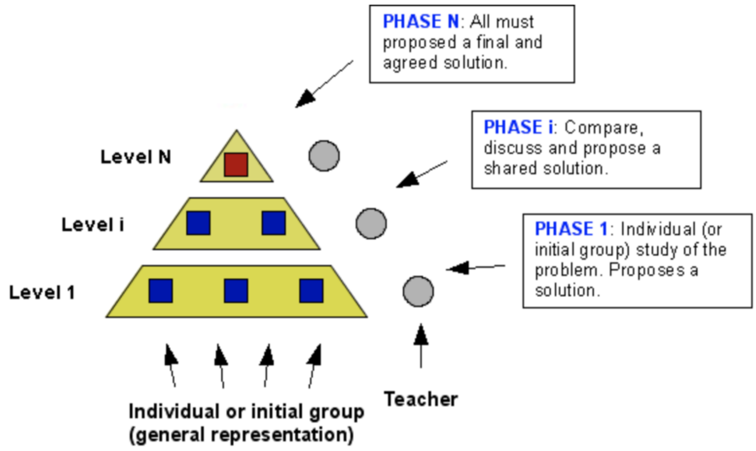


Figure 1: Pyramid CLFP (Hernández-Leo et al., 2006)

2.2 LM-GM Framework

According to Arnab et al., (2015). In gameful learning design one of the fundamental aspects consists in the translation of learning goals/practices into mechanical element of gameplay, “high-level pedagogical intents can be translated and implemented through low-level game mechanics”. To achieve this, first, they introduce the concept of Serious Game Mechanic (SGM) defined as the design decision that concretely realizes the transition of a learning practice/goal into a mechanical element of gameplay for the sole purpose of play and fun. SGMs act as the game elements/aspects linking pedagogical practices (represented through learning mechanics) to concrete game mechanics directly related to a player’s actions. Second, they propose the Learning Mechanics–Game Mechanics (LM-GM) model. This analytical model maps game mechanics and pedagogical elements that were abstracted from literature on game studies and learning theories.

The model helps to relate a set of standardized learning mechanics to another set of standard game mechanics. It allows for designers to investigate how the mechanics interact and to ensure that a system is grounded from a pedagogical and entertainment standpoint (Maarek et al., 2019), providing a concise means to relate pedagogy intentions and ludic elements within a player’s actions and gameplay (Arnab et al., 2015).

2.3 Gamification and Learning Analytics

Learning Analytics is the interpretation of a wide range of data produced by student's interaction in order to assess their academic progress, predict future performance and detect potential problems (Johnson et al., 2016). As stated by Gibson, (2017): data can "make it possible to gain highly detailed insight into student performance and their learning trajectories as required for personalizing and adapting curriculum as well as assessment".

In the particular context of game-based learning the use of learning analytics has been barely explored. By monitoring and analyzing gamification related data, experts can gain valuable insights and take corresponding actions towards goal achievement. Relevant data sources comprise user behavior data, user properties, and progression data (Heilbrunn et al., 2017).

Freire et al., (2016) explore the concept of Game Learning Analytics (GLA), its tools and technologies. Having data of what is happening while the user is playing is key to relating game-play with actual learning, and to move from only theory-based approaches to more data-driven or evidence-based approaches.

Additionally, other previous publications also explored the use of LA in gamification environments. For instance, the literature review performed by (Trinidad et al., 2018) shows the lack of tools that can provide gamification experts with real-time analytics from gamified systems, so experts can evaluate, improve and adapt their gamification strategies.

2.4 Related work

Previous studies have analyzed the effects of applying the LM-GM framework in the design of game-based learning applications. Callaghan et al. (2016), developed a case study where they applied the LM-GM in the design process of a serious game to teach electrical and electronic circuit theory. The objective of the game was to solve a series of circuit problems in stages, where the player explores each puzzle (behavioral momentum), tries to understand its structure and how to efficiently solve the problem using a simulate/response approach to observe, experiment and analyze circuit behavior under time constraints. The end of each level provides feedback to the player on their progress (score achieved), possible rewards (achievements) and competition (leader boards).

Baldeon et al. (2016), created 3DVirtualPC a serious game designed to develop computer literacy skills, such as the identification and analysis of concepts related to basic hardware components of a computer. To do so they made a mapping of the learning objectives, learning mechanics, game mechanics, and bloom taxonomy categories in order to decide the best implementation and in-game action.

However, to the best of our knowledge, none of the previous studies have addressed the use of the LM-GM framework in CLFPs, their pre-established features that can be gamified (e.g., phases, relevant expected student actions), and the gamification analytics required to monitor the students' actions supporting collaborative learning.

3 SELECTED IMPLEMENTATIONS OF PYRAMID CLFP SCRIPTS

In order to identify the relevant gameful characteristics presented in the Pyramid CLFP an analysis of learning - game mechanics and learning analytics used in diverse applications of the Pyramid CLFP was conducted. Three papers that report case studies that apply Pyramid were selected considering the following criteria: 1) they report explicit implementations of the Pyramid pattern, 2) they provide sufficient details of the learning design of the script, 3) the paper reports the use of learning analytics, 4) the study associated to the script is consolidated (i.e., vs. work in progress), 5) the three papers are written by different authors.

The application of these criteria led to the selection of the following journal papers:

Case 1. "Monitoring pattern-based CSCL scripts: a case study" by Rodríguez-Triana et al., (2011), which proposes a method to get an automated and higher level view about the evolution of the learning process structured by a CLFP to enable the monitoring of the collaboration.

Case 2. "Design of a Competitive and Collaborative Learning Strategy in a Communication Networks Course" by Regueras et al., (2011), presents a study using the tool QUESTOURnament that combines competitive and collaborative learning approaches in order to motivate students and improve their learning process.

Case 3. "Authoring and enactment of mobile pyramid-based collaborative learning activities in this paper" by Manathunga & Hernández-Leo, (2018), which proposes a particularization of the Pyramid CLFP to support flexible and scalable collaborative learning scenarios through the tool PyramidApp providing a web-based editor and an enactment environment accessible through web or mobile devices.

A summary of the main elements of each case can be found in Table 1.

Table 1: elements of each paper.

Sample	Learning design implementing the Pyramid	Additional tools	Observations
Case 1: 46 students of a third-year course (out of five) on "Network traffic and Management", s Engineering degree.	Two-level Pyramid CLFP. At level-1, groups of 2 participants attended a face-to-face lab session, students had to draw a preliminary version of a sequence diagram and write a report with a summary of the main decisions and issues. At level-2, groups joined in super-groups (composed of 4 groups). Each group had to review and provide feedback on the reports produced by their super-group mates; then, they had to discuss and produce a joint version of the diagram, and perform an oral presentation with a common version of the conclusions.	Students were provided with a shared board (Dabbleboard), and in order to explain, review and discuss, they had at their disposal shared documents and presentations (Google Documents and Google Presentations). Since these tools cannot be automatically integrated in Moodle the GLUE! Architecture (Alario-Hoyos et al., 2013) was used to integrate them into the VLE.	The authors reported some critical situations like for example that in a group, just one of the group members interacted with the resources or that in another group none of the participants had accessed the resources. Also, students complained about the lack of information about their mates' work, which forced them to connect and review their resources frequently to check whether there was any change.

Case 2: 36 students from an undergraduate communication networks course, part of the core curriculum of the three-year degree in an engineering program.	Two phases script: a 3 level pyramid and a competitive second phase. Initially at level-1 the students studied the class material individually, in the level-2 the students were grouped in pairs in order to prepare questions about one of the specific topics and add them to a wiki, in the level-3 the whole group participated in a discussion and selected the six best questions. In the competitive phase pairs students compete in order to answer the questions posed and then assessed by other students. Finally, the score for all of the work for each pair of students is calculated, taking into account both the score obtained for their answers and the assessment mark.	Moodle e-learning platform as a collaborative framework. Competitive active e-learning tool called QUESTOURnment, which is integrated into the Moodle platform. QUESTOURnament allows teachers to organize contests in both individual and group work environments. Each contest includes a set of intellectual challenges or questions that must be solved by the students within a certain time limit. The answers are rewarded by means of a variable scoring system, students can also submit new questions and assess the corresponding answers.	Throughout the experimental study, the teacher observed the learning process, reviewing both the proposed questions and the assessments of the answers, in order to guarantee the quality of the questions and the fairness of the assessments. In general terms, the students considered that it was easy to reach agreement. However, it was more difficult for them to pose questions and to assess the answers of their classmates, which is understandable since this is usually done by teachers and not by students.
Case 3: first-year undergraduate students (n = 194) taking an Introduction to Information and Communication Technologies, second-year students (n = 43) taking Network Protocols and Masters' students (n = 46) (several engineering programs) taking Research Methodology	Most of the PyramidApp rounds were conducted in f2f scenarios with different kinds of level configurations (all Pyramid activities having an individual level and one or two group levels). Two sessions (one with the first-year and another with the second-year) were enacted using the distance mode of the application. In a distance mode, students were receiving emails notifying about the activity progress, avoiding the need to be online all the time.	PyramidApp: a web-based scalable, dynamic collaborative learning application. that is used to orchestrate activities in which participants can express their individual solutions to a task followed by cumulative negotiations in increasingly larger groups (Pyramid levels) to select the most appropriate solution. The orchestration is done automatically considering the pedagogical constraints of the CLFP and a set of mechanisms that achieve flexibility in terms of flow dynamism and scalability.	Some students missed the initial submission phase due to either late access or ignored timing values instructed in the email. Several students could not submit options or rating timely as they were not paying attention to timer notifications. Some groups used the chat feature extensively while some did not Further investigations would be interesting for improving usage of discussion features.

4 GAME MECHANICS PRESENT IN PYRAMID CLFP SCRIPTS

For each paper an analysis of learning mechanics using the LM-GM framework was performed to identify the learning mechanics and the game mechanics that were applied in the scripted activity.

In case 1, the authors implemented a 2 level pyramid with face to face and virtual activities, the mapping of learning mechanics - game mechanics (Figure 3) shows that the script using a 2-level pyramid presents 8 of the 38 game mechanics. Students had to collaborate and cooperate with their peers to design a sequence diagram, then they had to provide feedback to the other groups, and in a process of communal discovery they created a joint version of the diagram.

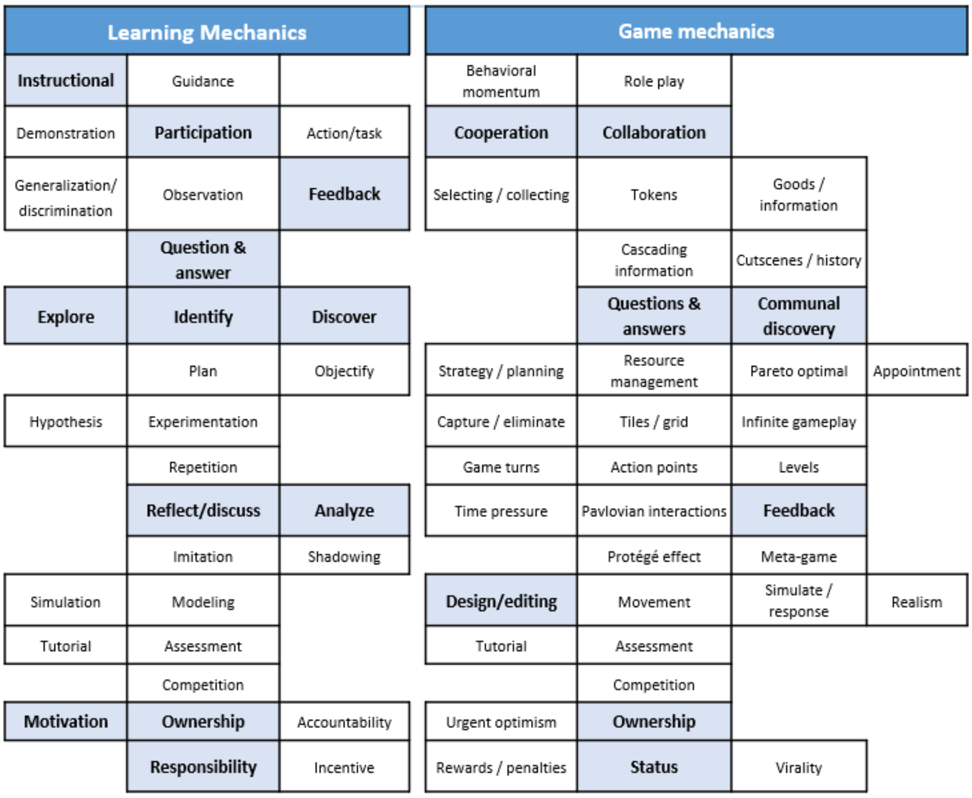


Figure 3: Learning and game mechanics in case 1 (elements in blue cells)

The use of a competition phase after the 3-level pyramid in the script used in case 2 added new learning and game mechanics, resulting in a total of 11 of the 38 mechanics included in the LM-LG framework (Figure 4). Students initially had to cooperate and collaborate to prepare questions about the studied topic and post them in a wiki, then in a communal discovery phase they had to select the six best questions, this created a sense of status and ownership between the students. In the competitive phase they had to compete to be the first to answer in a given period of time, having a second type of feedback which took into account both the score obtained for their answers and the assessment mark.

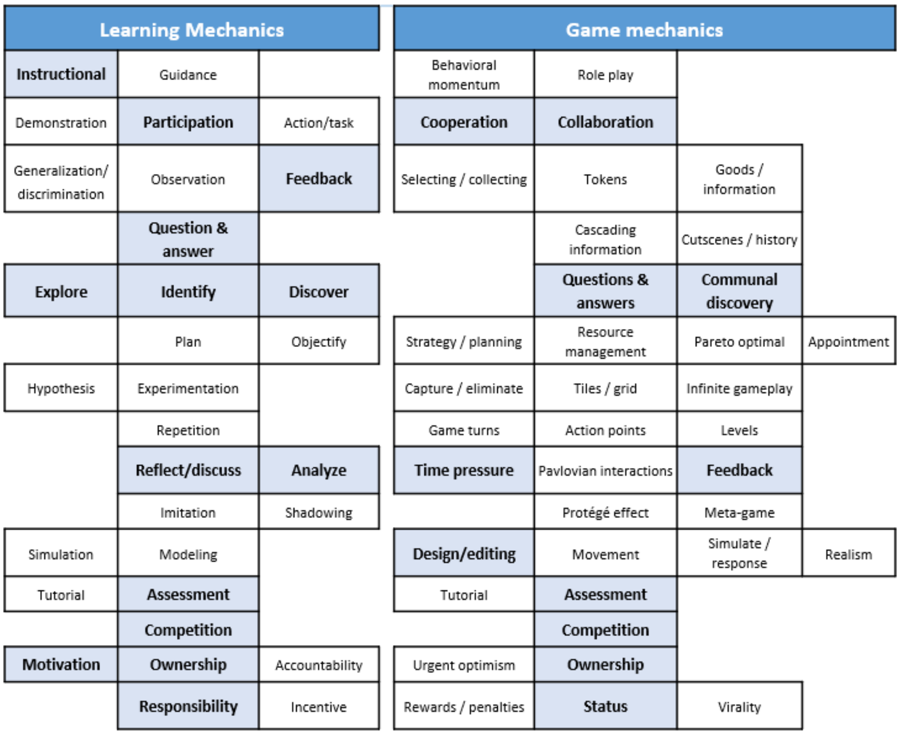


Figure 4: Learning and game mechanics in case 2 (elements in blue cells)

In case 3, the mapping is similar to the one obtained for case 1, with 9 of the 38 game mechanics (Figure 5). The addition of a level in the pyramid increased the amount of feedback received. Moreover, the use of a dedicated application that had a time limit to send the answers corresponds to the time pressure game mechanic).

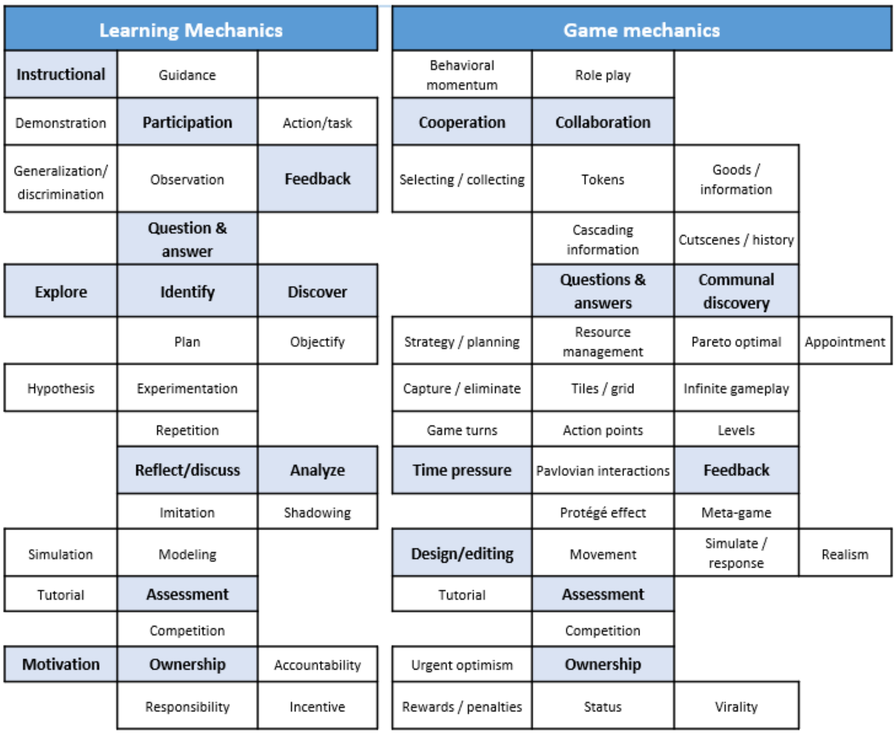


Figure 5: Learning and game mechanics in case 3 (elements in blue cells)

5 LEARNING ANALYTICS PRESENTED IN THE IMPLEMENTATION OF PYRAMID CLFP SCRIPTS

Each case had scripting mechanisms using CSCL tools that helped the enactment of the activities and allowed teachers and researchers to track the learning process. In this section we report an analysis of the learning analytics used in the cases, and how they relate to the different game mechanics.

In case 1, the authors retrieve and analyze the content of GLUE's logs, Moodle and Google docs event history (Table 2) in order to detect evidence of the key collaborative aspects that were previously estimated in the design of the activity, this analysis done during the enactment phase detected some potentially critical situations in the collaboration that were informed to the teachers in order to take preventive measures. The extracted data is related to the following game mechanics: cooperation, collaboration, design/editing and communal discovery.

Table 2: Learning Analytics data presented in case 1.

Tool	Source	Data
Google docs tools	Document revision history	User, date, time and document version
Moodle	Event history	Date, time, IP address, user name, action, resource used
GLUE!	Access History	Event logs user, date, time, resource accessed

Case 2 used a system integrated in the Moodle platform that allowed students and teachers to track the progress of the activity, showing the top ranked groups and their mark, students also could see the questions posted by others and their date (Table 3). In this case, the monitoring process was done manually by the teacher, however the gathered data along with a questionnaire that students had to answer at the end of the activity, was used to perform an analysis of the results comparing them with the ones of the subject in previous years.

Table 3: Learning Analytics data presented in case 2.

Tool	Source	Data
QUESTOURnament	history	User, date, time, assessment mark, final score, time used
Moodle	Event history	Date, time, IP address, user name, action, resource used
Wiki	Public data	Answer submitted, author, date
Survey	Answers	Work of the other groups, experience, perception of acquired learning

In case 2 the data gathered is mainly related to the competition and time pressure game mechanics, With the information from the questions posted on the wiki and the event history of QUESTOURnament the whole competition can be tracked, also the collaboration and cooperation game mechanics could be studied from the Moodle event history.

Case 3 used a dedicated application (PyramidApp) to help teachers in the design and enactment phase, during the activity students had to rank the different answers with a scale of 1 to 5 and could see which answers were promoted to the next level of the pyramid. The data gathered was used to track if the collaboration conditions were met (reach common goal, positive interdependence, coordination and communication, individual accountability, satisfaction).

Table 4: Learning Analytics data presented in case 3.

Tool	Source	Data
PyramidApp	logs	answers, chat logs, answer's score, time expended, ranking of the answers, users that submit
Survey	Answers	experience perception

In case 3, all the game mechanics presented in the script can be studied thanks to the use of a dedicated application that registers all the interactions during the enactment of the activity. The logs contain key information of the different participations during the script that can be used to develop a collaboration and cooperation measure. Also, the chat logs and the ranking of the answers can be used to track the game mechanics related to the status between the students and the ownership of the answer.

6 WHAT CAN BE PROPOSED TO EXTEND GAME MECHANICS AND THE USE OF LEARNING ANALYTICS FOR THE IMPLEMENTATION OF PYRAMID CLFP SCRIPTS?

The three case studies analyzed share common observations, for example: (1) in some groups the collaboration between participants was not as successful as expected (2) participants only interacted with some of the materials (3) participants did not submit their answers the given time. As shown in the LM-GM analysis, the Pyramid CLFP scripting implementation already incorporates a significant number of game mechanics. Yet, the analysis also shows that it is still possible to consider additional mechanics which has a potential to further improve the implementation of the pattern, e.g.:

- Use of a meta-game system that includes some of the classical elements of gamification (points, badges and leaderboards) linked with an in-course reward system can be a way to increase the level of interest and collaboration (Ortega-Arranz et al., 2018). The use of a competition system is optional but encouraged as reported in the second paper its application reported very good results. This meta-game will also generate analytics that can be used to track the level of engagement that the students have in order to help the teachers to take action

- Depending on the nature of the subject and the learning objectives, this can be augmented into a full role-playing gamification system, where students assume different roles (their progress might be also linked to an avatar) the success in the activity depends on the collaboration of the different roles with different characteristics. This game mechanic is very aligned with the "distribution of roles" mechanism present in several collaboration scripts, and it is known it can contribute to fostering positive interdependence, individual accountability and knowledge sharing (Kobbe et al., 2007), while at the same time creating more data variables to analyze and help teachers to track the achieving of the learning objectives.

7 CONCLUSIONS

Despite the increasing number of works in gamification, learning analytics and collaborative learning it is unclear to what extent there is or there could be a game-based perspective in the implementation of CSCL scripts, such as those structured according to the Pyramid Collaborative

Learning Flow Pattern. This paper analyses the integration of game mechanics and learning analytics in three reported case studies and shows that several game mechanics are present in the implementation of the Pyramid CLFP, such as collaboration, cooperation, competition, assessment, feedback, Communal discovery, status and ownership. Learning Analytics is used in the three cases, to track the learning process, determine the cooperation and collaboration levels and to inform the teachers about negative situations that could emerge during the activity. Game mechanics not considered in the cases include roleplay, rewards/penalties, history, levels and movement.

As future work, we plan to carry out co-design activities with teachers interested in using gamification in CLFPs in their teaching practice and evaluate the effects of their implementation with students. This will help us know (1) whether diverse game mechanics are seemed suitable for incorporation in learning designs involving playful CSCL scripts; (2) analyze whether the gamification strategies implemented actually contribute to further fostering positive interdependence, individual accountability and knowledge sharing, (3) determine which learning analytics would help practitioners in the design and delivery of the scripts.

ACKNOWLEDGEMENTS

This work has been partially funded by the EU Regional Development Fund and the National Research Agency of the Spanish Ministry of Science and Innovation under project grants TIN2017-85179-C3-3-R. D. Hernández-Leo acknowledges the support by ICREA under the ICREA Academia program.

REFERENCES

Alario-Hoyos, C., Bote-Lorenzo, M. L., Gómez-Sánchez, E., Asensio-Pérez, J. I., Vega-Gorgojo, G., & Ruiz-Calleja, A. (2013). GLUE!: An architecture for the integration of external tools in Virtual Learning Environments. *Computers and Education*, *60*(1), 122–137. https://doi.org/10.1016/j.compedu.2012.08.010

Arnab, S., Lim, T., Carvalho, M. B., Belloti, F., Freitas, S. de, Louchart, S., Suttie, N., Berta, R., & Gloria, A. De. (2015). Mapping Learning and Game Mechanics for Serious Games Analysis. *British Journal of Educational Technology*, *46*(2), 391–411. https://doi.org/10.1111/bjet.12113

Arnab, S., Lim, T., Carvalho, M. B., Bellotti, F., De Freitas, S., Louchart, S., Suttie, N., Berta, R., & De Gloria, A. (2015). Mapping learning and game mechanics for serious games analysis. *British Journal of Educational Technology*, *46*(2), 391–411. https://doi.org/10.1111/bjet.12113

Baldeon, J., Rodriguez, I., Puig, A., Gomez, D., & Grau, S. (2016). From learning to game mechanics: The design and the analysis of a serious game for computer literacy. *Iberian Conference on Information Systems and Technologies, CISTI*, *2016-July*. https://doi.org/10.1109/CISTI.2016.7521614

Callaghan, M. J., Mcshane, N., Eguiluz, A. G., Teilles, T., & Raspail, P. (2016). Practical application of the Learning Mechanics-Game Mechanics (LM-GM) framework for Serious Games analysis in engineering education. *Proceedings of 2016 13th International Conference on Remote Engineering and Virtual Instrumentation, REV 2016*, *February*, 391–395. https://doi.org/10.1109/REV.2016.7444510

Darejeh, A., & Salim, S. S. (2016). Gamification Solutions to Enhance Software User Engagement—A Systematic Review. *International Journal of Human-Computer Interaction*, *32*(8), 613–642. https://doi.org/10.1080/10447318.2016.1183330

Fluke, S. M., & Peterson, R. L. (2013). Positive behavior interventions & supports. *Strategy Brief.*

Student Engagement Project, University of Nebraska-Lincoln and the Nebraska Department of Education. https://k12engagement.unl.edu/Briefs/PBIS/PBIS 10-23-2013.pdf

Freire, M., Serrano-Laguna, Á., Iglesias, B. M., Martínez-Ortiz, I., Moreno-Ger, P., & Fernández-Manjón, B. (2016). Game Learning Analytics: Learning Analytics for Serious Games. In *Learning, Design, and Technology* (pp. 1–29). Springer International Publishing. https://doi.org/10.1007/978-3-319-17727-4_21-1

Gibson, D. (2017). Big Data in Higher Education: Research Methods and Analytics Supporting the Learning Journey. *Technology, Knowledge and Learning*, *22*(3), 237–241. https://doi.org/10.1007/s10758-017-9331-2

Heilbrunn, B., Herzig, P., & Schill, A. (2017). Gamification Analytics---Methods and Tools for Monitoring and Adapting Gamification Designs. In S. Stieglitz, C. Lattemann, S. Robra-Bissantz, R. Zarnekow, & T. Brockmann (Eds.), *Gamification: Using Game Elements in Serious Contexts* (pp. 31–47). Springer International Publishing. https://doi.org/10.1007/978-3-319-45557-0_3

Hernández-Leo, D., Aseniso-Pérez, J. I., Dimitriadis, Y., Bote-Lorenzo, M. L., Jorrín-Abellán, I. M., & Villasclaras-Fernández, E. D. (2005). Reusing IMS-LD Formalized Best Practices in Collaborative Learning Structuring. *Advanced Technology for Learning*, *2*(4), 223–232. https://doi.org/10.2316/journal.208.2005.4.208-0865

Hernández-Leo, Davinia, Asensio-Pérez, J. I., & Dimitriadis, Y. (2005). Computational representation of collaborative learning flow patterns using IMS learning design. *Educational Technology and Society*, *8*(4), 75–89. https://www.jstor.org/stable/jeductechsoci.8.4.75

Hernández-Leo, Davinia, Villasclaras-Fernández, E. D., Asensio-Pérez, J. I., Dimitriadis, Y., Jorrín-Abellán, I. M., Ruiz-Requies, I., & Rubia-Avi, B. (2006). COLLAGE: A collaborative Learning Design editor based on patterns. *Educational Technology and Society*, *9*(1), 58–71.

Jermann, P., Soller, A., & Lesgold, A. (2004). Computer Software Support for CSCL. In Strijbos, J.-W., P. A. Kirschner, & R. L. Martens (Eds.), *What We Know About CSCL* (pp. 141–166). Springer Netherlands. https://doi.org/10.1007/1-4020-7921-4_6

Johnson, D., Deterding, S., Kuhn, K. A., Staneva, A., Stoyanov, S., & Hides, L. (2016). Gamification for health and wellbeing: A systematic review of the literature. *Internet Interventions*, *6*, 89–106. https://doi.org/10.1016/j.invent.2016.10.002

Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., & Hall, C. (2016). Horizon Report - 2016 Higher Education Edition. In *NMC Horizon Report*. https://doi.org/ISBN 978-0-9968527-5-3

Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., & Fischer, F. (2007). Specifying computer-supported collaboration scripts. *International Journal of Computer-Supported Collaborative Learning*, *2*(2–3), 211–224. https://doi.org/10.1007/s11412-007-9014-4

Maarek, M., Louchart, S., McGregor, L., & McMenemy, R. (2019). Co-created design of a serious game investigation into developer-centred security. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11385 LNCS*, 221–231. https://doi.org/10.1007/978-3-030-11548-7_21

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. https://doi.org/10.1111/bjet.12588

Nah, F. F.-H., Zeng, Q., Telaprolu, V. R., Ayyappa, A. P., & Eschenbrenner, B. (2014). Gamification of Education: A Review of Literature. In F. F.-H. Nah (Ed.), *HCI in Business* (pp. 401–409). Springer International Publishing.

Ortega-Arranz, A., Kalz, M., & Martinez-Mones, A. (2018). Creating engaging experiences in MOOCs through in-course redeemable rewards. *IEEE Global Engineering Education Conference, EDUCON*, *2018-April*, 1875–1882. https://doi.org/10.1109/EDUCON.2018.8363464

Radkowitsch, A., Vogel, F., & Fischer, F. (2020). Good for learning, bad for motivation? A meta-analysis on the effects of computer-supported collaboration scripts. *International Journal of*

Computer-Supported Collaborative Learning. https://doi.org/10.1007/s11412-020-09316-4
Regueras, L. M., Verdú, E., Verdú, M. J., & De Castro, J. P. (2011). Design of a competitive and collaborative learning strategy in a communication networks course. *IEEE Transactions on Education*, *54*(2), 302–307. https://doi.org/10.1109/TE.2010.2053933
Rodríguez-Triana, M. J., Martínez-Monés, A., Asensio-Pérez, J. I., Jorrín-Abellán, I. M., & Dimitriadis, Y. (2011). Monitoring pattern-based CSCL scripts: A case study. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6964 LNCS*, 313–326. https://doi.org/10.1007/978-3-642-23985-4_25
Trinidad, M., Calderón, A., & Ruiz, M. (2018). A systematic literature review on the gamification monitoring phase: How SPI standards can contribute to gamification maturity. *Communications in Computer and Information Science*, *918*, 31–44. https://doi.org/10.1007/978-3-030-00623-5_3

The 3nd 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: teachers, students and researchers. The previous years workshops at LAK19 and LAK20 focused on reading behavior in higher education, and this year we will offer participants a new challenge that focuses on secondary school reading behavior. Due to data privacy and ownership, we will offer a synthetic dataset that has been generated by a model trained on actual data, a method that has been gaining wider attention in LAK community. As with previous years, 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, Synthetic Data

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., 2018; 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 et al., 2018; Flanagan et al., 2019; Flanagan et al., 2020), with 16, 14, and 17 participant submissions respectively. However, to-date the data challenges have targeted higher education settings.

This year we will offer participants a new challenge that focuses on secondary school reading behavior. Due to data privacy and ownership, we will offer a synthetic dataset that has been generated by a model trained on actual data, a method that has been gaining wider attention within the LAK community (Berg et al., 2016; Dorodchi et al., 2019). The use of synthetic data also broadens the possible scope of reading behavior data challenges to other areas of education that were previously limited. In the proposed workshop, we will offer a unique opportunity for participants to:

- Analyze **large-scale synthetic reading log data based on secondary school with performance-based labels for model training** from one subject.
- 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.
- Offer participants the opportunity to implement analysis trained on synthetic data in a real-world learning analytics dashboard.

This year we will provide two datasets: a large labeled training dataset and a smaller test dataset will be distributed. 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 periodically. A leaderboard will be provided with the best evaluation score that each participant has achieved to encourage competition between teams. Final data challenge results of prediction models will be confirmed by submission of prediction models for formal evaluation.

2 OBJECTIVES

While we welcome research questions from all participants, and we expect to emphasize the following topic which the organizers feel attention should be paid. Low retention and high failure rates are important problems in education (Villagrá-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 (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.

ACKNOWLEDGEMENTS

This work was partly supported by JSPS Grant-in-Aid for Scientific Research (B) 20H01722, JSPS Grant-in-Aid for Scientific Research (S) 16H06304 and NEDO JPNP20006 and JPNP18013.

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.

Berg, A. M., Mol, S. T., Kismihók, G., & Sclater, N. (2016). The role of a reference synthetic data generator within the field of learning analytics. *Journal of Learning Analytics*, *3*(1), 107-128.

Dorodchi, M., Al-Hossami, E., Benedict, A., & Demeter, E. (2019). Using Synthetic Data Generators to Promote Open Science in Higher Education Learning Analytics. In *2019 IEEE International Conference on Big Data (Big Data)* (pp. 4672-4675).

Flanagan, B., & Ogata, H. (2018). Learning Analytics Platform in Higher Education in Japan, Knowledge Management & E-Learning (KM&EL), 10(4), 469-484.

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*.

Flanagan, B., Majumdar, R., Shimada, A., & Ogata, H. (2020). The 2nd Workshop on Predicting Performance Based on the Analysis of Reading Behavior, *Companion Proceedings of the 10th International Conference on Learning Analytics and Knowledge*.

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.

Villagrá-Arnedo, C. J., Gallego-Durán, F. J., Llorens-Largo, F., Compañ-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.

Development of a Time Management Skill Support System Based on Learning Analytics

Hiroyuki Watanabe
Faculty of Arts and Science, Kyushu University
watanabe@artsci.kyushu-u.ac.jp

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: To achieve their learning goals, students need to make effective use of and manage their limited time. However, effective time management is not an easy task for many students. Notably, learning analytics is an effective approach that supports students with regard to their awareness of their time management. In this study, we report on the development of a system that can assist students in creating and managing their own study schedules by providing them with accurate learning data.

Keywords: learning skills, time management, learning analytics

1 INTRODUCTION

Learning skills are fundamental to academic achievement (Gettinger and Seibert, 2002), and include the abilities associated with acquiring, recording, organizing, synthesizing, remembering, and using information (Hoover and Patton, 1995). Additionally, successful students effectively use time-related learning skills such as time management, time planning, examination predicting and preparation, and reading and writing strategies (Zimmerman et al.,1996). These skills are related to self-regulated learning (SRL), which is an important component of the learner's learning process and environment. Zimmerman et al. (1996) stated that self-regulation is an effective perspective for achieving learning outcomes and learning goals through the learner's active involvement in the learning process-including metacognition, motivation, and behavior. Furthermore, recent advances in information and communication technology (ICT) have made it possible to trace learners' actual behaviors with a learning analytics approach. This awareness of SRL has also been shown to have a relationship with learning behaviors (Yamada et al., 2015, 2016a, 2017a, 2017b, 2018). Also, Chen et al. (2019a) indicated that cognitive learning behaviors in reading learning materials enhanced SRL awareness and learning performance. Yamada et al. (2016b) indicated that the awareness of SRL promotes effective that leads to the keeping of the submission deadline for an assignment. These studies show that learning skills can be linked to learning behavior. However, even if students have learning skills, it is difficult to implement them if they cannot manage their time. In this paper, we report on the development of the system to help students acquire time-management skills based on ordinal learning behavior data.

2 REVIEW OF PREVIOUS RESEARCH

An effect of time-management skills may have a positive impact on college performance (Britton and Tesser, 1991). Students felt that a study plan was beneficial to their learning (Goda et al., 2009) and time-related learning patterns have influences on their learning performance (Goda et al., 2015). Furthermore, learning time management is an important aspect of self-regulated learning (Wolters et al., 2017). Also, it has been demonstrated that factors such as time management and test anxiety are significantly related to academic performance (Talib and Sansgiry, 2012). Moreover, it has been reported that the benefits of time-management behaviors include a greater buffering effect on learning stress rather than leisure satisfaction activities (Misra and McKean, 2000). Thus, time management has been shown to be useful in alleviating learning stress and improving academic performance. Research on learning analytics has included not only psychological data but also learning logs and other data to contribute to the improvement of education and learning environments. For example, studies suggesting that reading learning materials promotes learning performance (Yin et al., 2019; Okubo et al., 2016), and out-of-class learning enhances learning performance (Shimada et al., 2015). Practical research on time-management has shown improvement in time management skills by tracing learning data using self-management tools in online learning (Tabuenca, 2014). In addition, data was collected in a computer course at a graduate school to detect time-management strategies from learning activities. The results showed that time-management was significantly related to grades (Uzir et al., 2020). The types of learning logs were identified to measure learning skills with learning analytics. This was done from the perspectives of basic learning skills including SRL, which is strongly related to learning skills (Watanabe et al., 2020a), and design of an application to help students manage their learning time (Watanabe et al., 2020b). Also, the learning analytics feature that was most accepted by students is the deadline reminder function, which is an important function for students' independent learning and time-management (Schumacher and Ifenthaler, 2018).

Thus, the learning analytics approach was effective in measuring time-management skills and the students also wanted to manage their own study time. Therefore, we developed a system (hereinafter referred to as the"MAI Helper") that allows them to manage their study time and control their learning activities. MAI means “Management and Active involvement of Learning”.

3 SYSTEM DEVELOPMENT AND FUNCTIONS

As for the functions of the MAI Helper, we developed the visualization of each graph, access time, schedule progress, reflection, comparisons between the students and their classmates, and the integration with BookRoll (BR) and Metaboard (MB). BR refers to electronic textbooks, which are allowed to be used to record the use of digital lecture materials such as slides and notes (Ogata et al., 2015, 2017a, 2017b). MB is a learning analytics dashboard that aims to provide the visualization of processes and behaviors based on the learning log data of the BR system (Chen et al., 2019b, 2020; Lu et al., 2020). Data from the BR and MB systems can be obtained and visualized on the MAI Helper and is displayed in Figure 1.

Below the MAI Helper logo is the list of courses registered in the schedule since the access date. The student's name, period, and subject are displayed on the right. On the left side of the screen, users can select the day and the week; accordingly, the deeper the color, the more the accesses to the site. If users select the "Schedule" tab, they can register the contents of each schedule. Regarding the "BookRoll," "Metaboard," and "Reflection" tabs on the left side, the darker the color, the more accesses there are. Students can also see their own access date and time, the class access date and time, and the number of people in the class by moving the cursor over it. On the right side, the schedule shows the access day and the students can check the detailed contents by moving the cursor in order to edit or delete items. In addition, there is a reminder function, and if students check the box when entering the schedule, they will be notified by e-mail at the desired date and time. If users select the Reflection tab, they can reflect on their learning and make plans for the next time. (Figure2)

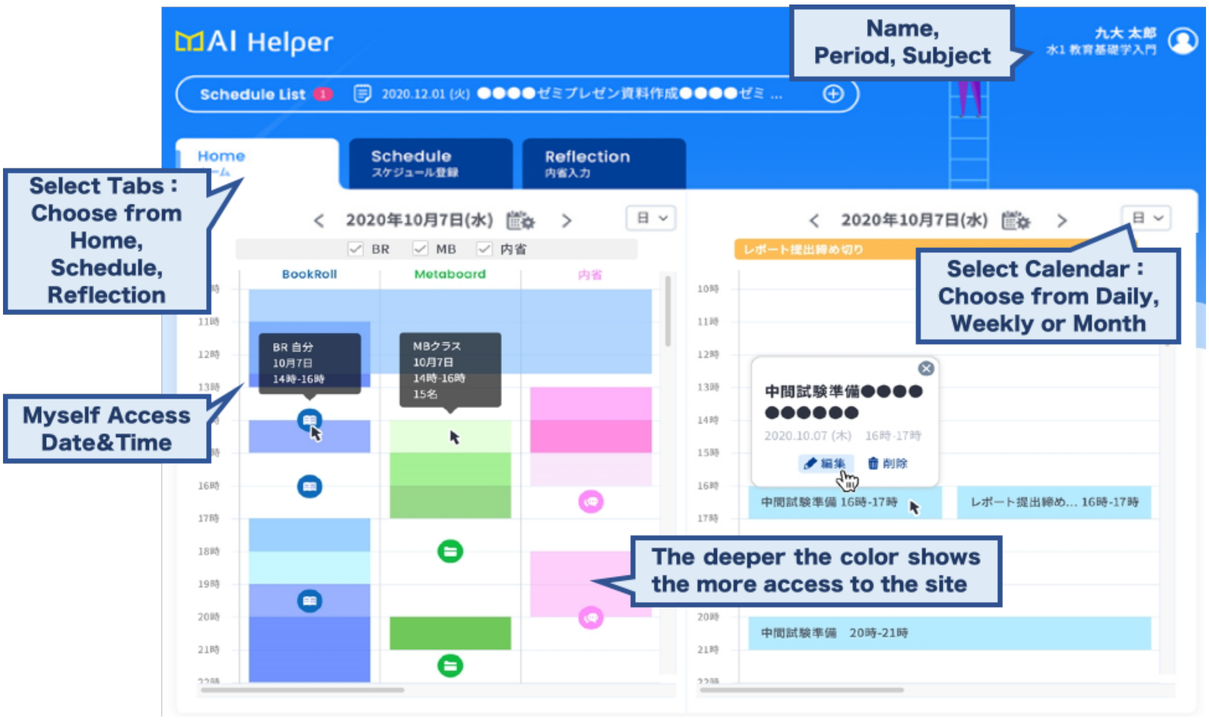


Figure 1: Top screen of MAI Helper

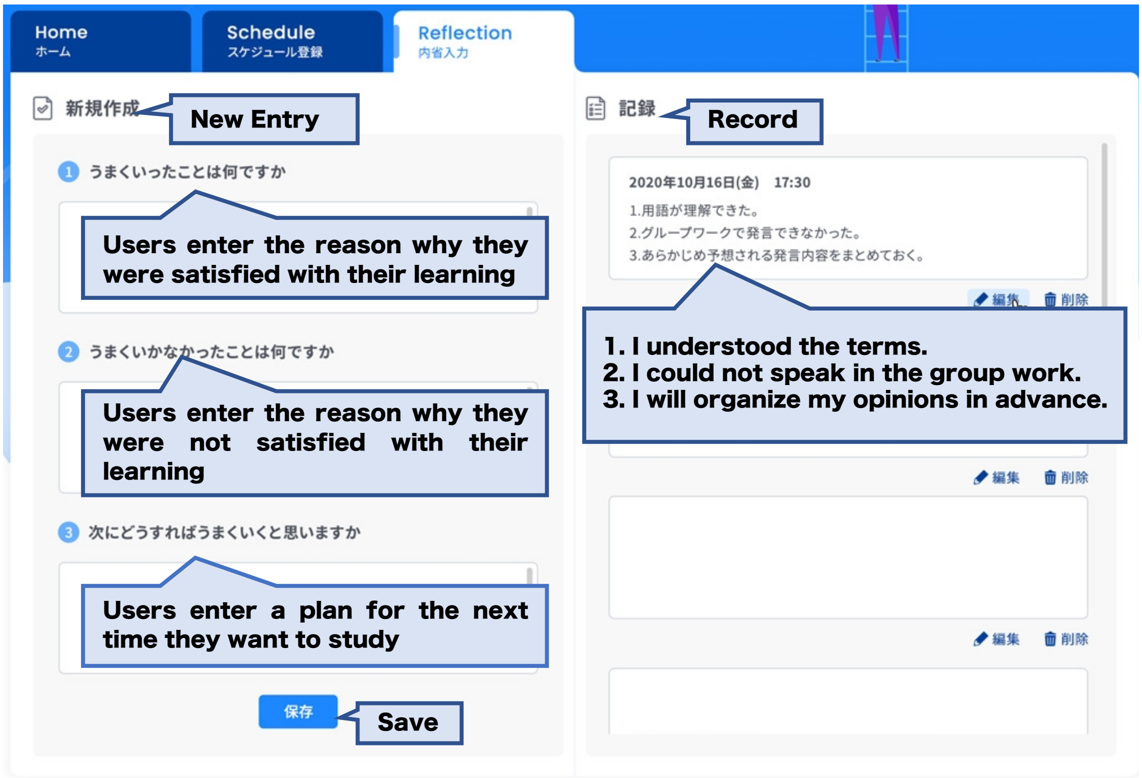


Figure 2: Reflection for students

Figure 3: Registration for teachers

In addition, the teachers-only functions include the registration of class information (Figure3) and the downloading of access logs (Table1). The background is colored in morning, noon, evening, and night so that the user can intuitively understand the access time.

Next, the specific configuration of the system involves a few steps. First, the students access the campus using a single sign-on system using a PC, tablet, etc. Then, they access the learning

Table 1: Student access log

No.	Physical name	Data type	Not Null	Default	Remarks
1	log_id	int(11) unsigned auto_increment	Yes (PK)		
2	user_id	varchar(191)	Yes		{Moodle user id}@{consumer key}
3	course_id	varchar(191)	Yes		{Moodle course ID}@{consumer key}
4	lti_properties	text	Yes		Keep the lti_properties values coming from metaboard
5	event_type	varchar(191)	Yes		[1:LOGIN]
6	created_at	timestamp on update CURRENT_TIMESTAMP	Yes	current_timestamp()	Creation date and time

management system (LMS), and after the Learning Tools Interoperability (LTI) linkage, they access the Learning Analytics Dashboard (LAD). The MAI Helper is then positioned as a kind of Learning dashboard. The database used by the MAI Helper is mysql (Ver 8.0.21).

4 FLOW OF USING THE SYSTEM

In this section, we describe how users engage with the system (Table 2). First, the students manage their learning time by using access times, schedules, and reflections. In addition to this, teachers can register class times, assignment submissions, discussions, tests, and so on. It also allows teachers to monitor the status of each student's activity. Until now, the data from learning analytics has been provided to faculty members but rarely to students. The reason for this is that even if the data is provided in its present state, it cannot be analyzed and used.

However, this time students can compare themselves with their classmates with regard to access time with the assistance of the MAI Helper, which has linked the MB, BR, and Schedule system. Moreover, they can use it for their own time-management.

Table 2: How to use the system

No.	Students	Teacher
1	Login to MAI Helper	Login to MAI Helper
2	Check today's schedule	Check today's schedule
3	Select a class subject	Edit your schedule
4	Select the day, week, and month.	Click on the Class Management Registration tab and enter the required information.
5	Select Home, Schedule or Reflection	Select the day, week, and month.
6	Select BR, MB or Reflection in Home to check your and your class' status	Check the access status of BR, MB, and reflection for the whole class
7	Check your progress, edit, and set reminders in Schedule	Check individual student access to BR, MB, and reflection
8	Review and edit your reflections	Check individual student reflection
9	Logging out of MAI Helper	Check the progress of an individual student's schedule
10		Log out of MAI Helper

Teachers can download the saved log data in comma-separated values (CSV) format. The main contents are access logs, course information, user information, students' schedule information (logs), reflection information(logs), and teacher-class information. In the case of MB, data related to the browsing of BR pages is also available.

5 WHAT WILL IT BE ABLE TO ANALYZE?

The analysis here is to find out to what extent the dependent variable (behavior) changes when the independent variable (environment) is manipulated by using the MAI Helper. In other words, what should we do to make good learning behavior habitual in students? This is what the MAI Helper is designed to support. The students compare their own access status with that of the class. If they want to know the details, they can access the MB and see which page of the text has been accessed the most. The MAI Helper, for example, is used to compare the time and number of accesses in the first two weeks with the time and number of accesses in the last two weeks, and to see how it has changed to see the adjustment in behavior.

Next, regarding reflection, we can confirm the change by looking at the number of words in the first week's statement and the number of words in the last week's statement. As for the schedule, we think it is possible to understand whether the students were able to implement the study plan according to the schedule by themselves. However, it may be difficult for students to perform these analyses. For this reason, it is necessary to provide explanations and guidance on analysis methods prior to implementation.

If we assume that the use of the MAI Helper changes behavior, we can grasp a certain trend by tracing these data. In addition, it would be more accurate to combine the trace data with the evaluation of the questionnaire for a more accurate analysis.

In contrast, the use of the MAI Helper from the teacher's side can be checked by the whole class or by individuals. For example, it is possible to send confirmation e-mails to students with low access to prevent dropouts. Furthermore, teachers can also explore the correlation between student behavior (number of accesses and reflections) and final grades. In this way, the MAI Helper can provide students with accurate learning data, which may help them manage and implement their own study schedules.

The MAI Helper stores learning logs on the database. Teachers and researchers can access the data for learning analytics in order to investigate the relationships among reading behaviors, learning plans, actual learning behaviors, and are the novelty of the MAI Helper is that it specializes in learning time management and are linked to MB and BR. Goda et al. (2015) found several learning patterns that enhance learning performance, but they did not analyze the various variables related to learning behaviors such as reflection. Additionally, Goda et al. (2015) indicated that procrastination leads to low learning performance, but several learners succeeded in gaining high performance, as Yamada et al. (2016b) indicated. It will be hopeful that researchers can find the threshold of successful active procrastination through learning analytics using the learning logs on the MAI Helper.

6 CONCLUSION AND FUTURE WORK

In this paper, we have developed a time-management system focused on student engagement by utilizing a learning analytics approach. This system is linked to BR and MB, and aims to help students manage their study time by themselves. MB does not visualize only BR logs. It is a platform for a variety of learning log visualization tools. In other words, a variety of learning log visualization tools can be placed on top of MB. MAI Helper is one of those tools.

The feature of this system is to make a learning plan while taking into account the actual learning time and the learning time of others. And the log of this system itself can be used to analyze the level of action taken to manage the study time.

However, in order to implement this system, it is necessary to provide guidance to students on how to use it. In the next phase, guidance for use will be included in Moodle, formative evaluation—including a user interface—will be conducted.

Future work includes the providing of measures for users to analyze, a system that can automatically sort reflective writing, and consideration for linking tasks and test schedules.

ACKNOWLEDGMENTS

This research is supported by the JST AIP Grant No. JPMJCR19U1 and JSPS Kakenhi JP19H01716, Japan. Moreover, we are grateful to Mr. Xue Wang Geng , Mr.Yufan Xu for checking this manuscript.

REFERENCES

Britton, B. K., & Tesser, A. (1991). Effects of time-management practices on college grades, *Journal of Educational Psychology*, 83(3), 405-410.

Chen, L., Goda, Y., & Yamada, M. (2019a). Factors investigation of learning behaviors affecting learning performance and self-regulated learning. In Proceedings of IEEE Teaching, Assessment and Learning for Engineering 2019, At: Yogyakarta, Indonesia, 676-681. DOI: 10.1109/TALE48000.2019.9225926

Chen, L., Lu, M., Goda, Y., & Yamada, M. (2019b). Design of learning analytics dashboard supporting metacognition. In Proceedings of 16th International Conference Cognition and Exploratory Learning in Digital Age (CELDA 2019), 175-182.

Chen, L., Lu, M., Goda, Y., Shimada, A., & Yamada, M.(2020) Factors of The Use of Learning Analytics Dashboard That Affect Metacognition. In Proceedings of 17th International Conference Cognition and Exploratory Learning in Digital Age (CELDA 2020), 295-302.

Gettinger, M., & Seibert, J.K. (2002). Contributions of Study Skills to Academic Competence. *School Psychology Review*, 31(3), 350-365. https://doi.org/10.1080/02796015.2002.12086160

Goda, Y., Matsuda, T., Yamada, M., Saito, Y., Kato, H., & Miyagawa, H. (2009). Ingenious attempts to develop self-regulated learning strategies with e-learning: Focusing on time-management skill and learning habit. In E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, 1265-1274. Association for the Advancement of Computing in Education (AACE).

Goda, Y., Yamada, M., Kato, H., Matsuda, T., Saito, Y., & Miyagawa, H. (2015). Procrastination and other learning behavioral types in e-learning and their relationship with learning outcomes. *Learning and Individual Differences*, 37, 72-80. https://doi.org/10.1016/j.lindif.2014.11.001

Hoover, J. J., & Patton, J. R. (1995). Teaching Students with Learning Problems to Use Study Skills: A Teacher's Guide. Austin, TX: Pro-Ed. https://doi.org/10.1177/002221948902200711

Lu, M., Chen, L., Goda, Y., Shimada, A., & Yamada, M. (2020, March). Development of a learning dashboard prototype supporting meta-cognition for students. In Proceedings of the 10th International Conference on Learning Analytics & Knowledge (LAK20) ,104-106.

Misra, R., & McKean, M.,(2000). College Students' Academic Stress and Its Relation to Their Anxiety, Time Management, and Leisure Satisfaction, *American Journal of Health Studies*, 16(1), 41-51.

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), 401-406.

Ogata, H., Oi, M., Mohri, K., Okubo, F., Shimada, A., Yamada, M., Wang,J., & Hirokawa, S. (2017a). Learning analytics for e-book-based educational big data in higher education. In Smart Sensors at the IoT Frontier, 327-350. Springer, Cham.

Ogata, H., Taniguchi, Y., Suehiro, D., Shimada, A., Oi, M., Okubo, F., Yamada, M. & Kojima, K. (2017b). M2B system: A digital learning platform for traditional classrooms in university. Practitioner Track Proceedings, 155-162.

Okubo, F., Hirokawa, S., Oi, M., Yin, C., Shimada, A., Kojima, K., Yamada, M., & Ogata, H. (2016). Learning Activity Features of High Performance Students. 1st International Workshop on Learning Analytics Across Physical and Digital Spaces, CrossLAK 2016 - Edinburgh, Scotland, United Kingdom, CEUR Workshop Proceedings, 1601 , 28-33.

Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior*, 78, 397-407. https://doi.org/10.1016/j.chb.2017.06.030

Shimada, A., Okubo, F., Yin, C., Kojima, K., Yamada, M., & Ogata, H. (2015). Informal learning behavior analysis using action logs and slide features in e-textbooks. In 2015 IEEE 15th International Conference on Advanced Learning Technologies, 116-117. IEEE.

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.

Talib, N. ,& Sansgiry, S. S. (2012). Determinants of Academic Performance of University Students, *Pakistan Journal of Psychological Research*, 265-278. https://doi.org/10.1080/03075079.2018.1496408

Uzir, N.A., Gašević, D., Matcha, W., Jovanović, J., & Pardo, A (2020). Analytics of time management strategies in a flipped classroom. *Journal of Computer Assisted Learning*, 36(1), 70-88 https://doi.org/10.1111/jcal.12392

Wolters, C. A., Won, S., & Hussain, M. (2017). Examining the relations of time management and procrastination within a model of self-regulated learning. Metacognition and learning, 12(3), 381-399.

Watanabe, H., Chen, L., Geng, X., Goda, Y. & Shimada, A. (2020a). Design of the Time Management Skills Acquisition System Using Learning Analytics. In Proceedings of 17th International Conference Cognition and Exploratory Learning in Digital Age (CELDA 2020), 67-73.

Watanabe, H., Chen, L., Shimada, A., & Yamada, M. (2020b). Toward Identification of Learning Skills from Learning Analytics Approach: A Research Review. In Proceedings of IEEE TALE 2020, An International Conference on Engineering, Technology and Education , 836-841.

Yamada, M., Goda, Y., Matsuda, T., Kato, H., & Miyagawa, H.(2015). The Relationship among Self-regulated Learning, Procrastination, and Learning Behaviors in Blended Learning Environments. In Proceedings of the 12th Cognition and Exploratory Learning in Digital Age 2015, 67-74

Yamada, M., Okubo, F., Oi, M., Shimada, A., Kojima, K., & Ogata, H. (2016a). Learning analytics in ubiquitous learning environments: Self-regulated learning perspective. In Proceedings of International Conference on Computers in Education 2016, 306-314

Yamada, M., Goda, Y., Matsuda, T., Saito, Y., Kato, H., & Miyagawa, H.(2016b), How does self-regulated learning relate to active procrastination and other learning behaviors? *Journal of Computing in Higher Education*, 28(3), 326-343

Yamada, M., Oi, M., & Konomi, S. (2017a). Are learning logs related to procrastination? From the viewpoint of self-regulated learning. In Proceedings of 14th International Conference on Cognition and Exploratory Learning in the Digital Age (CELDA 2017), 3-10. IADIS Press.

Yamada, M., Shimada, A., Okubo, F., Oi, M., Kojima, K. & Ogata, H. (2017b). Learning analytics of the relationships among self-regulated learning, learning behaviors, and learning performance, Research and Practice in Technology Enhanced Learning, 12, 13.

Yamada, M., Shimada, A., Oi, M., Taniguchi, Y., & Konomi, S. (2018). BR-Map: Concept Map System Using e-Book Logs. International Association for Development of the Information Society. (CELDA 2018), 248-254

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, 35(4-5), 313-322. DOI: 10.1080/10447318.2018.1543077

Zimmerman, B. J., Bonner, S., & Kovach, R. (1996). Developing self-regulated learners: Beyond achievement to self-efficacy. American Psychological Association, Washington DC, USA.

Automatic Classification of the Learning Pattern - Time-Series Clustering of Students' Reading Behaviors

Hiroyuki Kuromiya
Graduate School of Informatics, Kyoto University
Khiroyuki1993@gmail.com

ABSTRACT: Students have several learning patterns. Some prefer doing it early, but others prefer it last. To investigate students' learning patterns quantitatively, Learning Analytics (LA) is one of the best options that help instructors understand students' behaviors. In this paper, we propose an automated workflow of classifying students' learning patterns. Our paper revealed 1) the different learning patterns in the dataset, and 2) the relationship between students' progress patterns and performance. To apply time-series clustering, we used the *tslearn* package in Python. Changing the number of clusters from 1 to 10, we determined the number of clusters as 3 because the decrease of the distortion was saturated there. As a result of the adoption of time-series clustering, we obtained three kinds of learning patterns. Cluster 1 seems to be an early engagement group. Cluster 3 seems to be a continuous engagement group and Cluster 2 was the intermediate of them. According to the statistical test between cluster number and scores, we concluded that students in Cluster 3 tend to get a higher score than other learning patterns. It suggests effective intervention plans like a reminder at the end of the semester.

Keywords: Learning Analytics, Time-Series Analysis, Time-Series Clustering

1 INTRODUCTION

Students have several learning patterns. Some prefer doing it early, but others prefer it last. Some students prefer doing it steadily, but others prefer doing it intensively. To investigate students' learning patterns quantitatively, Learning Analytics (LA) is one of the best options that help instructors understand students' behaviors (Akcapinar et al., 2020b). It is also said that learning patterns can affect students' academic performance (Akcapinar et al., 2020a). So, investigating students' learning patterns is an important topic for predicting students' performance based on the reading behavior dataset. In this paper, we propose an automated workflow of classifying students' learning patterns with a time-series clustering approach. In this workshop, here are the contributions of our paper:

1. We revealed several different learning patterns in the dataset.

2. We explored the relationship between students' learning patterns and their performance.

The rest of our paper is structured as follows. In the next section, we will introduce time-series clustering models and their applications. The Method section includes the data processing flow including the way we adopt time-series clustering to the dataset. In the Result section, we show the clustering results and the relationship between the clustering and performance. In the Discussion section, we point out the possibility of effective intervention found from the analysis results.

2 TIME-SERIRES CLUSTERING

Time-series clustering is a clustering model that is specialized for time-series data. There are several clustering models and several distance measures in time-series data. For example, the most popular distance measure for time-series data is Dynamic Time Warping (Sakoe & Chiba, 1971). It enabled us to compute the similarity between two time-series data with different lengths and deal with the time-shift between them. KL distance (Dahlhaus, 1996) or Hidden Markov Models (Smyth, 1997) are also used as a similarity measure in the clustering.

For the applications of time-series clustering, the field of Finance is very popular. Kumar & Patel (2002) found the seasonality patterns in the retail study. Bagnail et al. (2003) revealed the personal income pattern from the financial data. On the other hand, there are a few examples of the application of time-series clustering in the educational field. Młynarska et al. (2016) analyzed Moodle activity data with time-series clustering and uncovered three meaningful engagement patterns – Procrastinators, Strugglers, and Experts. Hung et al. (2015) identified at-risk students with a time-series clustering approach.

In the learning analytics field, there are some examples such as learning pattern classification based on the sequence mining approach (Akcapinar et al., 2020a) or page transition analysis during the lecture time (Akcapinar et al., 2019). However, it is not attempted that exploring students' learning patterns with a time-series clustering approach using the whole semester dataset.

3 METHODOLOGY

To apply time-series clustering in this dataset, we used the *tslearn* package (Tavenard et. al., 2020) in Python. Here, we choose the "TimeSeriesKMeans" model because it is the simplest approach and fastest in the package. While the package offers three types of distance measures – "euclidean", "dtw", and "softdtw", we choose "euclidean" because we want to distinguish the time of the engagement. Here is our workflow for analyzing the dataset.

1. *Sampling*: For the convenience of the analysis, we randomly sampled 1,000 students from the original dataset. The dataset was obtained from the e-book reader BookRoll (Ogata et al., 2017; Flanagan et al., 2017).

2. *Feature Extraction*: We made a time-series vector for each student. The time-series vector was a one-hot vector, which is 1 if the student interacted with the e-book reader on the day otherwise 0. As we have 111 days in the dataset, the length of the vector was 111. This vector was accumulated over the period.

3. *Standardization*: Based on the extracted feature vector, we defined the 'achievement rate' for each student and each day. We divided each element of the vector by the sum of the vector for each student. Through this process, the final achievement will be 1 with everyone's data.

4. *Determination of the cluster number*: To determine the number of clusters, we conducted the Elbow plot (Figure 1). An Elbow plot is a plot where we plot the distortion (sum of the squared distance) of the clustering changing the number of clusters. Changing the number of clusters

from 1 to 10, we determined the number of clusters as 3 because the decrease of the distortion was saturated there.

Figure 1: Elbow Plot for Determine the Cluster Number

All the analysis process above was uploaded on the Gist (https://gist.github.com/Hiroyuki1993/d524f254e24239ef585c6accabf27de7).

4 RESULTS

4.1 Clustering Results

Figure 2 shows the clustering results. There were 330 (33%) students in cluster 1, 441 (44 %) students in cluster 2, and 229 (23 %) students in cluster 3. The red line represents the cluster center (typical progress chart) for each cluster.

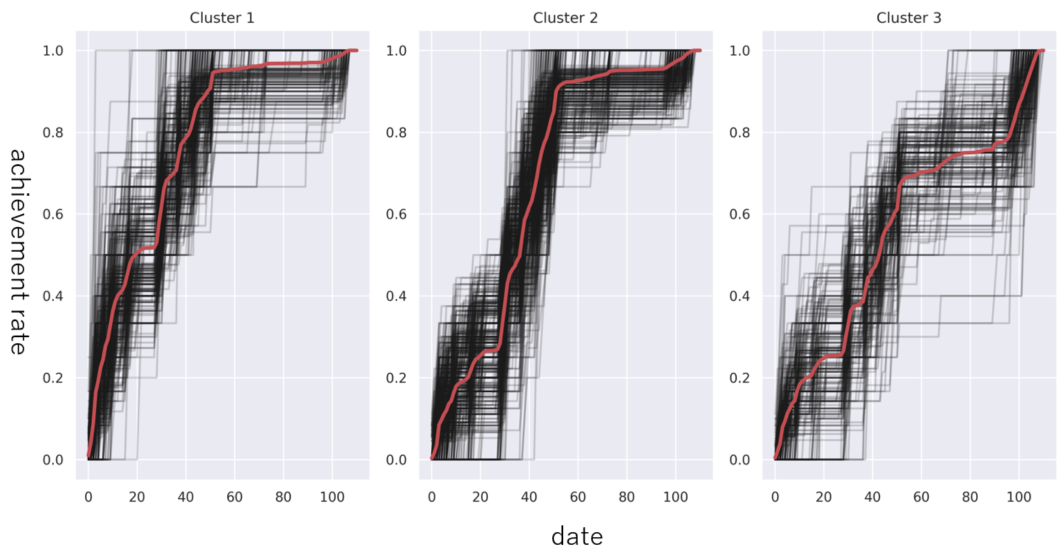


Figure 2: Time-Series Clustering of Students Learning Process

4.2 Relationship with Performance

We conducted a statistical test to check if there were any relationships between the cluster and the score. We used JASP (https://jasp-stats.org/) to conduct a statistical test (ANOVA). As the equality of the variance could not be observed ($p < .001$), we used Welch's homogeneity corrections. The result of ANOVA was significant ($p < .001$), and we conducted Post Hoc Comparisons between each cluster. The difference between clusters 1 and 3, and clusters 2 and 3 were significant ($p < .001$). The difference between clusters 1 and 2 was not significant ($p = 0.09$). Figure 3 shows the results of the analysis. The error bar stands for a 95 % confidence interval of the mean.

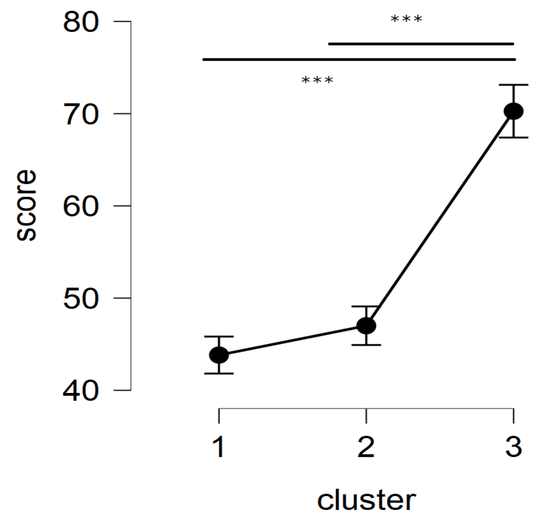


Figure 3: Average Score and Its Confidence Interval for Each Cluster

5 DISCUSSION

5.1 Clustering Results

As a result of the adoption of time-series clustering, we obtained three kinds of learning patterns. Cluster 1 seems to be an early engagement group. Students engaged hard during the first half of the semester but stop working hard after that. Students have achieved about 90 % of their engagement throughout the semester in the first half of the semester. On the other hand, students in cluster 3 studied hard at the end of the semester. The learning curve just before the end of the semester was the steepest of the three groups. Cluster 2 seems to be between Cluster 1 and 3.

5.2 Relationship with Performance

According to the statistical test between the cluster number and scores, we can conclude that students who engaged at the end of the semester tend to get a higher score than others. It suggests the engagement at the end of the semester may have affected their performance. It suggests effective intervention plans like a reminder at the end of the semester. It would be important how to motivate students at the end of the semester rather than the beginning of the semester.

5.3 Limitation

At the end of the paper, we pointed out two limitations in our study. The first is the standardization process. Our indicator – achievement rate – is defined as a relative measurement of students' engagement, so it may be wrong to compare different students' based on it. Second, we took a one-hot vector to represent students' learning patterns, but it would be more appropriate to use the bag-of-words (BoW) approach because it contains richer information about their learning. Classification based on the BoW vector would produce more reliable results for better intervention.

CONCLUSION

In this paper, we described an automated process of the clustering of students based on the time-series learning log data obtained from an e-book reader. Through the analysis, we found a significant relationship between learning patterns and performance. It will lead to better intervention for better performance.

REFERENCES

Akçapınar, G., Chen, M.-R. A., Majumdar, R., Flanagan, B., & Ogata, H. (2020a). Exploring student approaches to learning through sequence analysis of reading logs. Paper presented at the Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, Frankfurt, Germany.

Akçapınar, G., Hasnine, M. N., Majumdar, R., Chen, M. R. A., Flanagan, B., & Ogata, H. (2020b). Exploring temporal study patterns in eBook-based learning. Paper presented at the ICCE 2020 - 28th International Conference on Computers in Education, Proceedings.

Akçapınar, G., Hasnine, M. N., Majumdar, R., Flanagan, B., & Ogata, H. (2019). Using Learning Analytics to Detect Off-Task Reading Behaviors in Class. Paper presented at the 9th International Learning Analytics & Knowledge Conference (LAK'19), Tempe, Arizona.

Bagnall, A. J., Janacek, G. J., & Zhang, M. (2003). Clustering time series from mixture polynomial models with discretised data.

Dahlhaus, R. (1996). On the Kullback-Leibler information divergence of locally stationary processes. Stochastic processes and their applications, 62(1), 139-168.

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).

Hung, J. L., Wang, M. C., Wang, S., Abdelrasoul, M., Li, Y., & He, W. (2015). Identifying at-risk students for early interventions—A time-series clustering approach. IEEE Transactions on Emerging Topics in Computing, 5(1), 45-55.

Kumar, M., Patel, N. R., & Woo, J. (2002, July). Clustering seasonality patterns in the presence of errors. In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 557-563).

Młynarska, E., Greene, D., & Cunningham, P. (2016, September). Time series clustering of Moodle activity data. In 24th Irish Conference on Artificial Intelligence and Cognitive Science (AICS'16), University College Dublin, Dublin, Ireland, 20-21 September 2016.

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.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. IEEE transactions on acoustics, speech, and signal processing, 26(1), 43-49.

Smyth, P. (1997). Clustering sequences with hidden Markov models. In Advances in neural information processing systems (pp. 648-654).

Tavenard, R., Faouzi, J., Vandewiele, G., Divo, F., Androz, G., Holtz, C., ... & Woods, E. (2020). Tslearn, a machine learning toolkit for time series data. Journal of Machine Learning Research, 21(118), 1-6.

Addressing Drop-Out Rates in Higher Education

Agathe Merceron[1], María Jesús Rodríguez-Triana[2], Irene-Angelica Chounta[3], Juan I. Asensio-Pérez[4], Geoffray Bonnin[5], François Bouchet[6], Anne Boyer[5], Armelle Brun[5], Mohamed Amine Chatti[7], Yannis Dimitriadis[4], Vanda Luengo[6], Petra Sauer[1]

Beuth University of Applied Sciences Berlin, Germany[1], Tallinn University, Estonia[2], University of Tartu, Estonia[3], Universidad de Valladolid, Spain[4], Université de Lorraine, CNRS, Loria, France[5], Sorbonne Université, France[6], University of Duisburg-Essen, Germany[7]
merceron@beuth-hochschule.de, mjrt@tlu.ee, chounta@ut.ee, juaase@tel.uva.es, bonnin@loria.fr, francois.bouchet@lip6.fr, anne.boyer@univ-lorraine.fr, armelle.brun@loria.fr, mohamed.chatti@uni-due.de, yannis@tel.uva.es, vanda.luengo@lip6.fr, sauer@beuth-hochschule.de

ABSTRACT: This proposal describes the goal and activities of the second edition of the half-day symposium on Addressing Drop-Out Rates in Higher Education (ADORE 2021). As initiated in 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 institutional initiatives 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, often 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, which clearly illustrates the "impact LA solutions have on learning" (LAK 2021 conference theme[1]). This symposium will bring together established research practices from various contexts (that is, different countries, different academic institutions and different domains), extending the 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 threefold:

- To maintain and extend the community of stakeholders already started in ADORE 2020;
- 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;
- 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.

To achieve this aim, participants will be asked to read accepted submissions in advance. During the workshop, instead of a traditional presentation, accepted submissions will be discussed thoroughly in groups. 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

[1] https://www.solaresearch.org/events/lak/lak21/

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.

2.2 Proposed Schedule and Duration

This symposium is planned as a half-day event. We propose the following schedule:

- 9:00am-9:15am: Welcome, introduction, and goal of the workshop.
- 9:15am-9:30am: Attendees present themselves shortly.
- 9:30am-10:30am: Group discussion of accepted submissions.
- 10:30am-10:45am: Break (15 minutes).
- 10:45am-11:15am: Group discussion of accepted submissions.
- 11:15am-11:45am: Discussion: shaping best practices and building a knowledge base
- 11:45am-12:00 pm: 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 enable group discussions of the accepted papers, posters and demos. 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 LAK2021 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 seven different academic institutions from four countries, and at least the attendance of representatives from these institutions is guaranteed.

2.6 Required equipment

As the event is purely online, it would be convenient to have access to an online meeting platform that enables group work with breakout rooms.

3 WORKSHOP GOALS

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

ACKNOWLEDGMENT

Work by Irene-Angelica Chounta was funded by the Estonian Research Council (PUT grant PSG286)

REFERENCES

Petra Sauer François Bouchet, Vanda Luengo, Geoffray Bonnin, Anne Boyer, Armelle Brun, Mohamed Amine Chatti, Irene-Angelica Chounta, María Jesús Rodríguer-Triana, Kairit Tammet, Agathe Merceron. (2020). Addressing Drop-Out Rates in Higher Education (ADORE 2020). Companion

Proceedings of the 10th International Conference on Learning Analytics & Knowledge (LAK'20), 740-743.

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.educause.edu/-/media/files/library/2019/4/2019horizonreport.pdf

DiSEA: Analysing Success and Dropout in Online-Degrees

Monique Janneck
Technische Hochschule Lübeck
monique.janneck@th-luebeck.de

Agathe Merceron, Petra Sauer
Beuth University of Applied Sciences Berlin
{merceron, sauer}@beuth-hochschule.de

ABSTRACT: Although several research works show that students at risk of dropping out of a course or a study program can be predicted with relatively high accuracy, this information has so far often not been accessible to course directors, teachers, or students. The DiSEA project aims to research this issue and close this gap in the context of online-degrees. Building on previous research, machine learning methods will be used to identify risk and success factors. The overall aim is to develop an integrated model to predict success in digital study programs and derive recommendations and interventions for course design, student counseling, and student self-reflection. A user-centered design involving all stakeholders will be followed.

Keywords: Online-degree, dropout, data analysis, dashboard, user-centered design

1 INTRODUCTION

High dropout rates are a problem for students as well as for universities. Young people experience a failure or even a serious cut in their career path. Economically, this represents a waste of educational resources. Relatively high rates of dropouts are leading universities to take innovative measures to actively address this problem. These include improved analytical studies of existing data.

For this purpose, models are developed, trained, and evaluated using machine learning methods (Aulck et al. 2019). The available data play an important role in the quality of the model (Schneider et al. 2019). These models can be used in early warning systems to identify students at risk earlier and provide more targeted advice.

One example of this is the early warning system *FragSte* (Berens et al. 2019, Schneider et al. 2019). *FragSte* only uses data that every university in Germany collects anyway. *FragSte* was evaluated at a state university and a private university of applied sciences. It was shown that the performance of this early warning system improves considerably as soon as data on academic performance can be accessed. Both FragSte and other (international) research projects (for example in Dekker et al.2009, Aulck et al. 2019) examine data from traditional study formats, i.e. face-to-face study programs with predominantly "classic" first-year students who start their studies in early adulthood or after leaving school.

In addition to traditional study formats, which still make up the majority of study offers, digital study offers are becoming increasingly important and will probably come into even greater focus as a result of the Corona pandemic. Digital study programs take place entirely or predominantly as online or distance learning offerings. Presently, in terms of the characteristics of students, they differ significantly from students in traditional study programs: The target group of digital study formats are typically people who are already full-time in professional life and would like to gain new or further qualifications or who, due to their personal life situation - e.g. bringing up children, caring for relatives, illnesses - would find it difficult to realize a study program in face-to-face teaching and appreciate the flexibility of online offers. Accordingly, these people also show different study behavior: They often study part-time, which goes hand in hand with a - planned - significantly longer duration of studies and often also lower prioritization of studies. The demands on students' self-regulation and motivation skills are significantly higher: although online formats offer a lot of flexibility to adapt studies to personal life situations, they also require a high degree of discipline and organizational skills (Minks et al. 2011). This means that the reasons for dropping out of modules or studies in digital study programs might also differ from traditional study formats. For working people, for example, the semester is often more difficult to plan; unforeseen work-related demands can prevent a module from being successfully completed. In some cases, students underestimate the amount of work required at the beginning of their studies, or the compatibility of studies and work turns out to be more difficult than previously assumed. In some cases, the obtention of a formal degree can become less important in the course of studies - for example, if the acquisition of qualifications can also be proven by intermediate certificates or employers already reward the competencies acquired during studies through better pay or promotion, even if a formal degree is still not completed. The higher demands for independence and self-organization are also not successfully mastered by all students. All these aspects mean that dropout rates are generally higher in online formats (Diaz 2002, Beard & Harper 2002, Baker et al. 2015).

Up to now, research on academic success and dropout has focused primarily on traditional study formats. Digital study formats are only marginally addressed. For this reason, it is questionable whether and how these results can be generalized for digital study formats.

The paper addresses these challenges and presents as research in progress the DiSEA project. DiSEA will focus on analysing and identifying factors for success/failure and dropout, especially in digital study formats, investigating the transferability of previous research results to digital study formats.

The paper is organized as follows. Section 2 reviews related literature. Section 3 describes the goals and research questions of the DiSEA project. In Section 4, we focus on the challenges with the involvement of different types of stakeholders and show how a Human-Centered-Design approach can meet them. Section 5 concludes the paper.

2 RELATED WORKS

Different kinds of data have been used to predict success and dropout in university programs. The first kind of data comes from the registrar's office. This data quite often includes demographic features as well as the academic performance of students, such as enrollments and marks. Different studies show that demographic data or data collected before enrolment are less meaningful for predicting academic success than data on academic performance. Such a finding is reported in (Aulck et al. 2019) that analyzed data from the academic administration of an American university of more than 66,000 students who began their studies between 1998 and 2010. Berens et al. (2019) analyzed data from two universities in Germany, one with 23,000 students and 90 different undergraduate programs and one with about 6,700 students and 26 undergraduate programs, and obtained the same result. This conclusion was also reached earlier by Dekker et al. (2009), although in this study only data from the first year of study of 648 students from a single degree program were considered.

Studies differ with about the machine learning models they use to predict dropout. For example, Dekker et al. (2009) achieved the best results with decision trees, Aulck et al. (2019) with logistic regression, and Berens et al. (2019) with the ensemble method AdaBoost.

Berens et al. (2019) and Aulck et al. (2019) use models that can be applied across universities. Therefore, academic performance is described with so-called "global features" that are not specific to a study program (such as the number of courses passed, average grade, or the number of courses taken). Manrique et al. (2019) compared the performance of models with "global features" and "local features" (i.e. program-specific performance such as grades in certain courses) and were able to show that better prediction results can be achieved with local attributes than with global attributes. The findings are inconsistent concerning the number of semesters considered: Berens et al. (2019) achieve better results the more semesters they consider for the prediction, while the study by Manrique et al. (2019) showed the opposite picture. One possible explanation is that the less good results of Manrique et al. (2019) are due to the small amount of data, as there are fewer dropout data in higher semesters.

In addition to studies that look at academic success globally, there are attempts to develop early warning systems for individual courses. In these attempts, the data is quite often the interactions stored by the learning management system, such as frequency of use, completion of assignments and their assessment, or by some specific learning software. “Course Signals", for example, aims to predict which students are at risk of failing a course (Arnold & Pistilli 2012) in a classical context, while Baneres et al. (2020) also examine specific risk factors in individual courses in an online-context and report high accuracy in predicting completion a course. Baker et al. (2015) and Kuzilek et al. (2015) were able to show for individual online courses that students who engage with the online materials early and regularly in the course and complete corresponding course assignments are more likely to pass the course successfully. Van Goidsenhoven et al. (2020) show that it is possible to achieve accurate predictive models of student success based on log data from an online course and that the models provide reasons for student success. Akçapınar et al. (2019) were also able to show for an e-book-based course that students who achieved better course outcomes had interacted more frequently and intensively with the online materials. These findings suggest that it is promising to look at data from learning management systems in digital study formats.

What form a subsequent intervention should take is not easy to decide. One early warning system described by Jayaprakash et al. (2014) is used in such a way that teachers contact students identified as being at risk. However, in addition to an improvement in student outcomes, this paper also reports a significant increase in course dropouts, which was not intended. Note, that this study took place in a classical context of a college education. By contrast, Baneres et al. (2020) report a slight reduction of dropout in the context of online-degrees: dashboards for students and teachers have been developed to warn students (and their teachers) who are at risk of failing a course. An experiment shows that dropout was slightly reduced for students who consented to the experiment, though it is not clear whether the warning-system itself is the reason for this reduction. Several dashboards have been designed and integrated into the learning management system or virtual learning environment to support students' reflection on their learning. Presently, there are not many studies measuring their usage and their impact. In a pilot study, de Quincy et al. (2019) reported that about 25% of the students view the dashboard weekly. More research to understand their usage, impact, and usefulness is needed.

Following the works of (Aulck et al. 2019) and (Berens et al. 2019), we intend to use primary data on academic performance to predict whether students are at risk of dropping out of their degree. We will investigate whether local or global features work the best in our context. However, it is not clear whether the features considered in the literature can be overtaken as is in our context where most of the students study part-time. Therefore, feature engineering will be investigated further also to provide understandable explanations of the prediction to students. Predicting "dropping out of the degree" will be complemented by "predicting dropping out of the course" making use of the data stored by the learning management system as done in (Baneres et al. 2020). It is an innovative aspect of this project to investigate how both predictions can be combined and conveyed constructively to students.

3 GOALS AND RESEARCH QUESTIONS

The DiSEA project aims to identify risk factors for dropout as well as success factors for online study programs. On the one hand, the transferability of previous research results to digital study formats will be investigated. On the other hand, analyses of data on learning behaviour will be combined with analyses of academic data.

For this purpose, the extensive experience and data from the university network "Virtual University of Applied Sciences", or VFH (Virtuelle Fachhochschule, https://www.vfh.de/), will be used. The VFH network was founded in 2001 as part of an extensive German research programme. Currently, 13 higher education institutions from several federal states and one from Switzerland belong to the network; students come from all over Germany, some from abroad. The VFH currently offers 12 joint accredited Bachelor's and Master's degree programmes as online degree programmes.

The VFH study programs rely on a common learning management system (Moodle), thus providing extensive user data.

Our project addresses the following main research questions:

1. *Generalizability of prior findings on dropout risk factors to digital study formats*: We will analyze if and to what extent current results and prediction models from traditional face-to-face study programs can be applied to online degrees. Furthermore, we will investigate which specific factors need to be taken into account to predict success and failure in digital study formats.

2. *Analysis of data on learner behavior*. As prior research shows, models predicting study success are relatively weak in the first semesters, when data on academic achievements is scarce. However, the introductory phase can be decisive for later success. In digital study formats, data from learning management systems (LMS) is available from the onset, providing insights into learning behavior (e.g. frequency and intervals of use). Therefore, we will analyze Moodle user data to enhance models predicting study success.

3. *Using dashboards to enhance learners' self-reflection.* In online study programs, students' self-regulatory capacities are crucial for success, as students need to structure and organize their learning activities themselves to a much higher extent than students in traditional face-to-face programs. In this regard, it is essential that students receive feedback on their learning activities to recognize and reflect problematic habits and the need for change. Learning management systems may use so-called dashboards to visualize user data, learning activities, deadlines, and assignments, etc. (e.g. Brandenburger et al. 2019, Constapel et al. 2019). In our project, we will develop a dashboard providing learning and study-related data to enhance students' self-regulatory competencies. These dashboards will be evaluated in selected courses to analyze their impact on learning behavior and success as well as user acceptance. Providing learning analytics dashboards for students is a rather new direction in learning analytics research, as teachers have long been the predominant target group (cf. Schwendimann et al. 2016).

The overall aim is to develop an integrated model to predict success in digital study programs. We expect that from these findings, we can derive additional recommendations e.g. for course design and student counseling.

4 CHALLENGES REGARDING THE INVOLVEMENT OF STAKEHOLDERS

To successfully conduct our research activities and enable practical changes various stakeholders need to be involved, first of all, students and lecturers, but also program managers, heads of department, student counselors, data security officers, etc. Especially students' active involvement in designing research activities is crucial. Prior studies show that students accept data analyses if they are well informed and convinced of potential benefits (cf. Ifenthaler & Schumacher 2016, Slade et al. 2019).

In our project, we will design our research and development activities in a *Human-Centered-Design* approach. Essential stakeholders will be included in all phases of the project:

- In the *requirements analysis* phase we will conduct workshops and interviews to include stakeholders' views. As pointed out in Martinez-Maldonado et al. (2016), a challenge of this first stage is the identification of "possible new and radical features that can be offered by the data to address stakeholder needs, but where the stakeholders may not realize this". For example, many students are not aware and even might not believe that dropout from a degree can be predicted with pretty high accuracy at the end of the first semester of study as shown in various works (Berens et al. 2019, Wagner et al. 2020).

- We will use a *rapid prototyping* approach to discuss and test conceptual ideas, especially with students and lecturers. That way they will be able to test design ideas and give concrete feedback and suggestions for improvements.
- Our *evaluation* concept includes qualitative as well as quantitative methods, e.g. interviews, usability tests and questionnaires. That way, we will be able to combine in-depth feedback and more subjective views with a large-scale quantitative evaluation.
- We will develop an integrated *Learning Analytics concept*, including best practice collections and recommendations for lecturers, student counselors, and program and course designers. This is aimed at providing hands-on advice on how to incorporate our research findings into everyday practice.

An overall challenge of this *Human-Centered-Design* approach is to motivate various stakeholders to be part of the adventure. As far as students are concerned, de Quincey et al. (2019) describe an interesting approach involving four student-ambassadors who reach out to teams. A pilot study described in Brun et al. (2020) reports the involvement of more than 300 students in the design of dashboards without describing in detail how they reached out to students. Similarly, Rodriguez-Triana et al. (2018) describe the involvement of teachers and students in the design of dashboards for teachers. We aim to make participation as easy and rewarding as possible, for example by offering incentives for participation in online surveys, but also including interviews and focus groups as part of course achievements. The last point is particularly important in the context of online-degrees where students are more mature, work full-time, have family, and, therefore, a tight schedule. Furthermore, a number of the degree programs offered are connected to Information Technology (IT) and Computer Science. They include courses like “Human-Computer Interaction”, “Data Base”, “IT-Law”, “Algorithms”, “Artificial Intelligence”, “User Experience”, or “User-Centered Design”. Thus, group work and discussion involving topics of the project like “user interfaces”, “data”, “data privacy”, “trust” or “explanations of models”, to name a few, allow for a fruitful combination of teaching and research. Such a procedure would also ease the participation of different teachers in the project.

However, reliance on volunteers might also present a methodical problem, as volunteers might be more motivated than students in general, making it harder to generalize the results. Therefore, it will be crucial to motivate critical students who do not wish to release their user data to participate in other forms of data collection, e.g. interviews and questionnaires.

As a first step, data collection and analysis will be strictly voluntary. Students will be informed in detail about what and how data is used for analysis. There will be an easy opt-out possibility, only data from students who consented to data analysis will be included in our research. This approach will allow us to build trust and confidence with all stakeholders of the project. Depending on the outcomes of this approach, data collection and analysis might be broadened.

5 CONCLUSION

This article presents the DiSEA project. A key objective of the project is to provide personalized advice via early warning systems to those students who are at risk of dropping out. The focus of the project is on digital study formats. The project will use the extensive experience and data of the university network "Virtual University of Applied Sciences". The analysis of academic data from the participating universities, combined with data on learning activities from the Moodle learning management system available at the VFH, is intended to provide early indications of the risk of dropping out. A major challenge in the project will be how to communicate the analysis results via suitable dashboards to the students concerned and the other people involved, such as student counseling, program management, lecturers, and how an early warning system can be used in practical applications. Challenges in this context are to communicate results in an explainable and comprehensible way. The trust of students, in particular, must be awakened through early participation in a user-centered or even participatory design.

REFERENCES

Akçapınar, G., Hasnine, M.H., Majumdar, R., Flanagan, B., & Ogata, H. (2019). Developing an early-warning system for spotting at-risk students by using eBook interaction logs. *Smart Learning Environments 6*:4, https://doi.org/10.1186/s40561-019-0083-4

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*, April 29-May 02, 2012, Vancouver, British Columbia, Canada.

Aulck, L. S., Nambi, D., Velagapudi, N., Blumenstock, J. E., & West, J. (2019). Mining University Registrar Records to Predict First-Year Undergraduate Attrition. In *Proceedings of the 12th International Conference on Educational Data Mining* (Montreal, Canada, July 2-4). EDM 2019, 9-18. http://educationaldatamining.org/edm2019/proceedings/

Baker, R.S., Lindrum, D., Lindrum, M. J., & Perkowski, D. (2015). Analyzing Early At-Risk Factors in Higher Education E-Learning Courses. I*nternational Conference on Educational Data Mining (EDM),* Madrid, Spain.

Baneres, D., Rodriguez, M. E., Guerrero-Roldán, A.E. & Karadeniz, A. (2020). An Early Warning System to Detect At-risk Students in Online Higher Education. *Applied Sciences 2020*, 10(13), 4427; doi:10.3390/app10134427 https://doi.org/10.3390/app10134427

Beard, L.A., & Harper, C. (2002). Student Perceptions of Online versus on Campus Instruction. *Education, 122* (4).

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. *Journal of Educational Data Mining, 11*(3), 1–41. https://doi.org/10.5281/zenodo.3594771.

Brandenburger, J., & Janneck, M. (i.V.). A teamwork tool to support group work in online-based higher education: Exploring user experience and the use of support mechanisms by students. *Submitted*.

Brandenburger, J., Constapel, M., Hellbrück, H., & Janneck, M. (2019). Analysis of Types, Positioning and Appearance of Visualizations in Online Teaching Environments to Improve Learning Experiences. *Proceedings of AHFE 2019*, Springer.

Brun, A., Gras, B., Merceron, A.: Building Confidence in Learning Analytics Solutions: Two Complementary Pilot Studies In: Adoption of Data Analytics in *Higher Education Learning and Teaching*, edited by D. Ifenthaler & D. Gibson, Springer, ISBN: 978-3-030-47392-1, 2020. pp. 285 -303.

Constapel, M., Doberstein, D., Hoppe, H. U., & Hellbrück, H. (2019). IKARion: Enhancing a learning platform with intelligent feedback to improve team collaboration and interaction in small groups. *Proceedings of the 18th International Conference on Information Technology Based Higher Education and Training*.

de Quincey, Ed., Briggs, C., Kyriacou, T. & Waller, R. (2019). Student Centred Design of a Learning Analytics System. Proceedings of Proceedings of the 9th International Conference on Learning Analytics & Knowledge, Association for Computing Machinery, New York, NY, USA, https://doi.org/10.1145/3303772.3303793.

Diaz, D.P. (2002) Online Drop Rates Revisited. *The Technology Source*, May/June 2002.

Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. *Education Technology Research and Development, 64,* 923–938. https://doi.org/10.1007/s11423-016-9477-y

Jayaprakash, S.M., Moody, E.W., Lauría, E.JM., Regan, J.R., and Baron, J.D. 2014. Early alert of academically at-risk students: An open source analytics initiative. Journal of Learning Analytics 1, 1 (2014), 6–47.

Martinez-Maldonado, R., Pardo, A., Mirriahi, N., Yacef, K., Kay, J., & Clayphan, A. (2016). LATUX: an Iterative Workflow for Designing, Validating and Deploying Learning Analytics Visualisations. *Journal of Learning Analytics*, *2*(3), 9-39. https://doi.org/10.18608/jla.2015.23.3

Minks, K.-H., Netz, N., & Völk, D. (2011). *Berufsbegleitende und duale Studienangebote in Deutschland: Status quo und Perspektiven*. HIS: Forum Hochschule 11/2011.

Schneider, K., Berens, J., & Burghoff, J. (2019). Drohende Studienabbrüche durch Frühwarnsysteme erkennen: Welche Informationen sind relevant? *Zeitschrift für Erziehungswissenschaft, 22,* 1121–1146. https://doi.org/10.1007/s11618-019-00912-1.

Schwendimann, B., Rodriguez-Triana, M., Vozniuk, A., Prieto, L., Boroujeni, M., Holzer, A., Gillet, D., Dillenbourg. P. (2016). Perceiving learning at a glance: A systematic literature reviewof learning dashboard research. IEEE Transactions on Learning Technologies.

Slade, S., Prinsloo, P., & Khalil, M. (2019). Learning analytics at the intersections of student trust, disclosure and benefit. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge LAK19*, March 2019, 235–244. https://doi.org/10.1145/3303772.3303796

Rodriguez-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. *Proceedings of Proceedings of the 8th International Conference on Learning Analytics &*

Knowledge, Association for Computing Machinery, New York, NY, USA, pp.417-426, https://doi.org/10.1145/3170358.3170364

Wagner, K., Merceron, A. & Sauer, P., (2020). Accuracy of a Cross-Program Model for Dropout Prediction in Higher Education. In Companion Proceedings of the 10th Learning Analytics and Knowledge Conference (LAK'20), p. 744-749. Workshop on Addressing Dropout Rates in Higher Education, Frankfurt am Main, Germany, 2020. https://www.solaresearch.org/wp-content/uploads/2020/06/LAK20_Companion_Proceedings.pdf

An overview of analytics for curriculum understanding and optimization in Higher Education

Liyanachchi Mahesha Harshani De Silva[1], María Jesús Rodríguez-Triana[1], Irene-Angelica Chounta[2], Gerti Pishtari[1]
Tallinn University[1], University of Tartu[2]
mahesha@tlu.ee

ABSTRACT: The use of Curriculum Analytics (CA) helps teachers, learners, as well as other institutional stakeholders to make evidence-based decisions at the program level to improve student success and reduce dropouts. This paper presents the first insights of a systematic literature review on Curriculum Analytics at Higher Education Institutions to determine 1) existing CA solutions proposed in the literature for Higher Education; 2) how such solutions have been used; and 3) the maturity of those solutions. Based on the review's findings, the paper presents limitations of the studies and proposes recommendations for future research in this field.

Keywords: Curriculum Analytics; Learning Analytics; Higher Education; Systematic Literature Review

1 INTRODUCTION

Higher Education Institutions (HEIs) -including universities, colleges, professional and teacher-training schools, junior colleges, and institutes of technology- are in pressure to evolve their strategies to increase student success and completion rates (Tinto, 2005). Many different factors may influence student success and dropout at the personal and institutional level, e.g. student choices, educational goals, personal reasons, the curriculum quality or the institutional support (Tinto, 2005).

Among the different strategies to overcome these issues, especially during the last decade, HEIs have used Learning Analytics (LA) solutions in order to offer different insights related to learning and teaching activities. While many LA solutions have focused on the improvement of teacher and learning strategies, improvements at the curriculum level are also necessary to address problems that go beyond the classroom context (Gottipati & Shankararaman, 2018). To target this need, Curriculum Analytics (CA), a subfield of LA, can be used to raise awareness and inform curriculum-related decision-making among program managers and directors (Ochoa, 2016).

While many reviews have been done in the field of LA in the last few years (e.g., Sclater et al., 2016; Mangaroska & Giannakos, 2018; Vieira et al., 2018; Larrabee et al., 2019; Romero & Ventura, 2020; Ifenthaler & Yau, 2020), none of them have focused on the area of CA. In fact, little is known about how CA tools facilitate the improving curriculum (Hilliger et al., 2020). Thus, a review of existing CA solutions would help to better understand the current state and existing gaps. Therefore, the purpose of this article is threefold: 1) to identify existing CA solutions proposed in the literature for higher education; 2) to understand how they have been used; and 3) to assess the maturity of those solutions.

2 RELATED WORK ON ANALYTICS FOR LEARNING DESIGN

Within the field of LA, several authors have highlighted the importance of connecting learning design and analytics (e.g., Lockyer et al., 2013; Rodríguez-Triana et al., 2015; Bakharia et al., 2016). While this connection could have benefits for both sides, in this paper we pay special attention to how analytics solutions can inform learning design decisions (Hernández-Leo et al., 2019), more concretely in relation to the curriculum. Evenmore, while the term *curriculum* could refer to lessons, seminars, workshops, courses and degree programs (Fraser and Bosanquet, 2006), we will focus on course and degree programs.

As reported in the review done from Mangaroska & Giannakos (2018) on LA for learning design, most of the papers remained at the learning activity level or focused on analysed teaching practices not specifically connected with the curriculum. On the contrary, the number of papers related to curriculum-related decision-making are very scarce. The recent LA review conducted by Ifenthaler & Yau (2020) shows in general how existing LA solutions facilitate study success in HEIs. However, the data-based decision to improve study success at the different course and program levels is not explicitly stated in the current reviews (Greer et al., 2016). Further, there is a paucity of evidence on how students' success depends on different curriculum aspects (Hilliger et al., 2020). Thus, there is a need for further understanding of the state of art in CA, especially, raising awareness about the contributions done so far, the stakeholders involved, and the maturity of the solutions. In summary, it is necessary to understand how CA stands regarding the rest of the LA field.

3 METHODOLOGY

As justified in the previous sections, the purpose of this paper is to address the following research questions: *1) What are the existing CA solutions proposed in the literature for HE settings?; 2) How have CA solutions been used?;* and *3) What is the level of the maturity of those solutions?* To answer these questions, we have carried out a systematic literature review following the guidelines provided by Kitchenham and Charters (2007).

As part of the review design, we selected six popular databases related to technology-enhanced learning and LA, namely: ACM Digital-Library[1], IEEE XPLORE[2], ERIC[3], ScienceDirect[4], Wiley[5]. These databases have been selected based on the past systematic reviews in this field (e.g., Schwendimann et al., 2016; Mangaroska & Giannakos, 2018; Ifenthaler and Yau, 2020). To identify the papers related to our research goals, we looked for papers where the core contribution was about curriculum analytics, or use a data mining, institutional, learning or educational analytics solution to improve the curriculum or curricula. Thus, we used the following query: *"Curriculum Analytics" OR "Curricula Analytics" OR (("Institutional analytics" OR "Learning Analytics" OR "educational analytics" OR "data mining") AND ("curriculum" OR "curricula"))*.

[1] http:// dl.acm.org/dl.cf
[2] http://ieeexplore.ieee.org/ Xplore/home.jsp
[3] https://eric.ed.gov
[4] http://www.sciencedirect.com
[5] https://onlinelibrary.wiley.com

While conducting the review, we queried the databases in between January 24th to 26th 2021 and yielded 4418 entries in total (see Figure 1). Since each database used different search engines and filtering criteria, we ran a script to automatically select those papers where the query terms appeared either in the title, abstract or keywords in order to have a homogenous dataset. After removing duplicates, we assessed all papers to comply with the inclusion and exclusion criteria presented in Table 1.

Figure 1: Stages of the systematic literature review

Table 1: Inclusion and exclusion criteria

Criteria	Description
Core contribution	The core contribution was about curriculum or curricula analytics, or use a data mining, institutional, learning or educational analytics solution to improve the curriculum or curricula.
Type of curriculum	The review covered studies at the course or program level. Thus, studies focused only on lessons were excluded.
Context	The article targeted HE.
Publication type	Short paper contributions such as conference posters and abstract-only publications were excluded.
Accessibility	The full text was available.
Versioning	In case of several publications about the same contribution, the most "mature" was taken into consideration for the review.
Language	Publications were written in English.

Out of 375 papers, 48 satisfied the inclusion and exclusion criteria, and were part of the systematic review (see Appendix A). For each paper, we extracted the following aspects:

- *Type of contribution*: including type of publication (e.g., reports, conference or journal papers) and type of research contribution (e.g., models, tools, frameworks, etc).
- *How the contribution was used:* including target stakeholders of the analysis results (e.g., students, teachers, curriculum designers or researchers), the granularity of the curriculum (e.g., course or program), the key purpose of the study (understanding vs. optimizing), supported curriculum aspects, as well as the data sources, data gathering and analysis techniques.

- *Maturity of the contribution*: including stakeholders involved, type of evaluation (e.g., proof of concept, expert evaluation, authentic case study, etc.), and focus of the evaluation (e.g., usability, accuracy, adoption, …).

The outcome of the coding process is summarized and can be consulted as additional material[6]. The following section reports on the first results from the review.

4 RESULTS

Out of 48 reviewed papers, 50% were journal and 50% conference papers. When reporting the results, we used aggregated numbers since there are studies which have more than one way of supporting curriculum, data sources, data gathering techniques, data analysis techniques etc. This section reports on the results following the research questions.

RQ1) What are the existing CA solutions proposed in the literature for HE settings?

We grouped papers based on the type of research contribution tagged by the authors. Out of 48 papers, 28% of the papers proposed *processes* to assess entire course materials, evaluate curriculum coherence. *Models* were the core contribution of 24% papers, including linear regression models for predicting the placement of students, explicit learner models, models for students results prediction, and planning course registration model. *Frameworks* followed the list of more frequent contributions (20%), structuring course-adapted student LA, critical dimensions of LA, or curriculum assessment. The 17% of the papers presented *tools* which provide a visual based analysis to discover the strengths and weaknesses of the curriculum and help the curriculum committee for continuous curriculum improvement. Next, 9% of the papers focused on *methods.,* e.g. to study the levels of curriculum importance and student satisfaction. Finally, 2% of the papers presented *architectures* for areas covered in the system such as architecture for game-based learning. This architecture helps curriculum designers to understand the impact of such a learning method to the curriculum compared to the traditional teaching-learning process.

RQ2) How have CA solutions been used?

LA Purpose. Attending to SOLAR's LA definition[7], LA may have two purposes: understanding or optimising learning and the environments. In this review, 46 papers (95.8%) focused on understanding the curriculum, and 2 papers (8.3%) went one step further taking actions to improve it. Some of those steps are adopting the curriculum to the dynamic changes in the industry and helping students identify the optimal curricula based on the students' educational history.

Curriculum support. In terms of the *granularity* of the curriculum, most of the papers (41, 85.4%) referred to programs while 6 (12.5%) of them focused on courses (only one paper the granularity of

[6] Paper codification: https://tinyurl.com/y3dd2md2

[7] SOLAR's LA definition: *LA is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs*). https://www.solaresearch.org/about/what-is-learning-analytics/

the curriculum was not stated (2.1%). In terms of the kind of proposed solution, Figure 2 provides an overview of the main aspects of the curriculum that the reviewed papers tried to address.

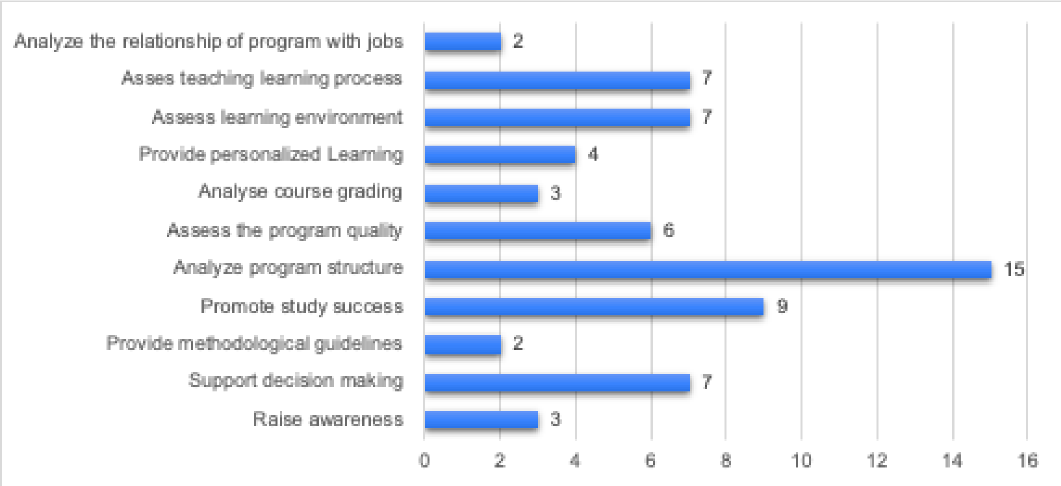


Figure 2: Types of curriculum support

For instance, some of these papers aimed at *assessing* course materials, the coherence of the program, the student preferences and academic needs on the curriculum, the curriculum's alignment with the industry expectations, the student learning processes, or to what extent the students have achieved the needed competencies based on the current curriculum.

Other papers tried to *identify new ways of improving teaching practices* (eg., looking at curriculum-level factors that affect retention and student outcomes, the difficulty level of the curriculum from the student perspectives, academic gaps and overlaps in the curriculum), to *identify good practices among students* (e.g., best study path students must traverse to acquire higher results), or to identify what resources are necessary for the improvements, to offer a better curriculum and a more personalised learning experience. Figure 3 shows the relationship between the CA solutions (RQ1) and type of curriculum support. According to the results, we can see that most of the studies have provided different processes, models, tools and methods for analyzing program structure.

Target users. The intended target users of the selected studies were curriculum designers 27 (56.3%), students 14 (29.2%), administrators 13 (27.1%), teachers 10 (20.8%), program curators 6 (12.5%), and researchers 1 (2.1%). As Figure 4 shows, it should be noticed that there were several studies which addressed different users in their proposals. The total size of each stakeholder group is represented on the left barplot. The bottom plot represents every possible intersection, and their occurrence is shown on the top barplot.

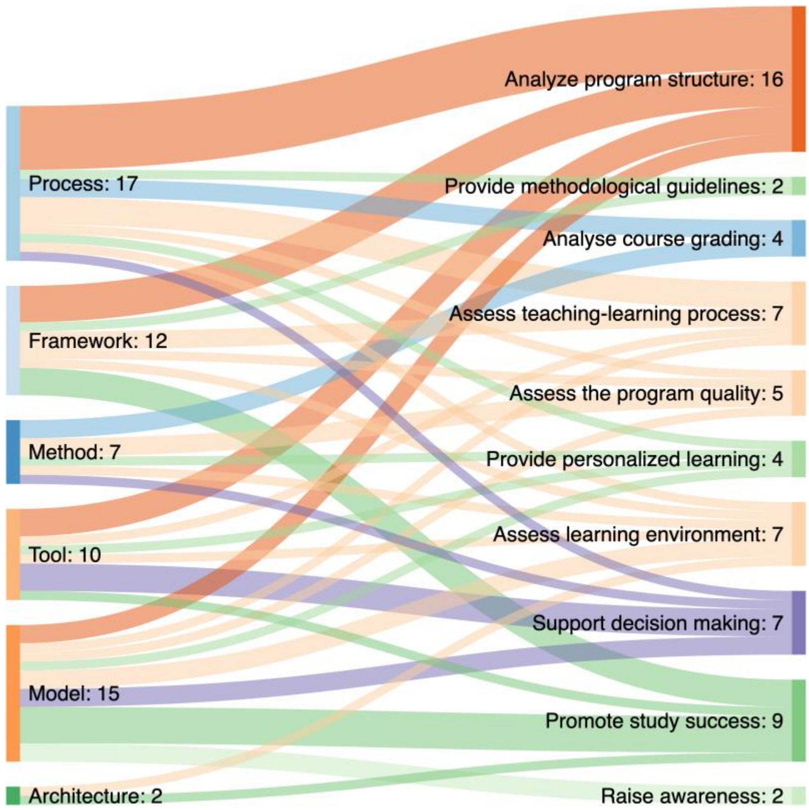


Figure 3: Relationship between types of curriculum support and CA solutions

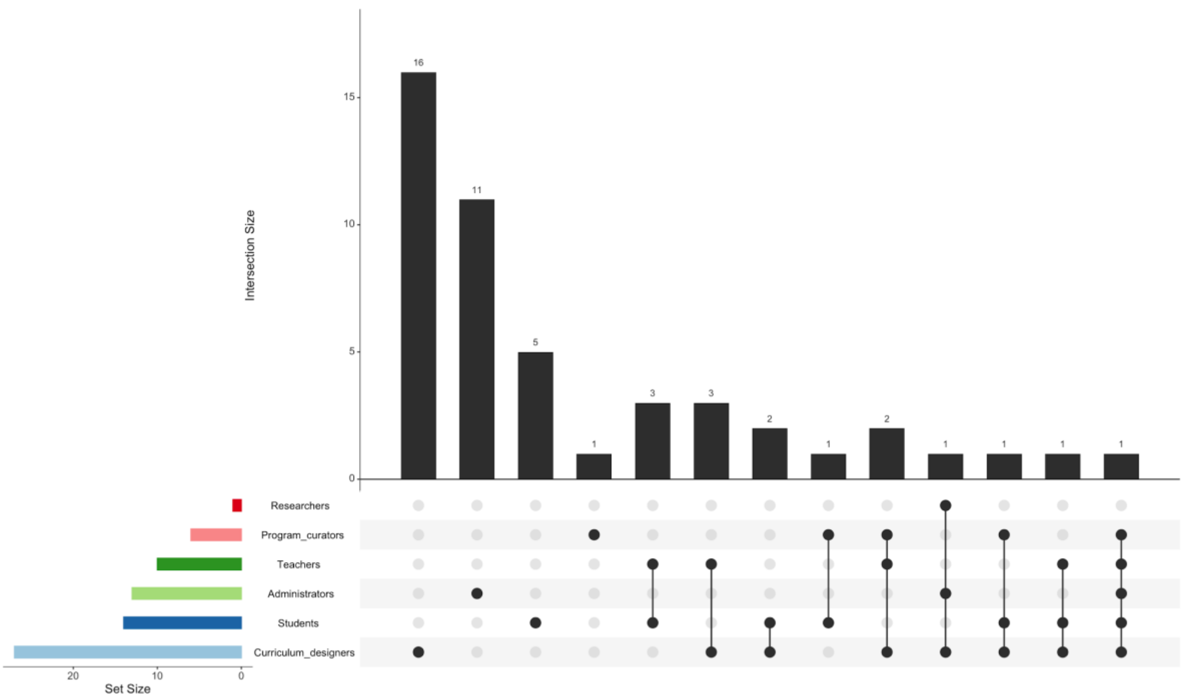


Figure 4: The target users of the paper contributions

Data sources. The analysis of data sources used in the studies shows that interestingly, 11 (22.9%) papers did not use or report the data sources of their studies. Among those mentioning the data sources, even if not all the data sources were reported, most of them used already-existing data from the learning ecosystem: 16 (33.3%) from institutional management system, 6 (12.5%) from learning management systems (e.g., Moodle or Blackboard), 6 (12.5%) other learning tools (e.g., chat or student feedback tools, YZU virtual classroom, clinical log. In addition, 3 (6.3%) papers used the university website as a data source and other 3 (6.3%) papers extracted data from non-academic websites (e.g., Job bank, The library of congress, or LinkedIn). Finally, it is noteworthy that 12 papers (25%) collected ad-hoc data directly from the stakeholders.

Data gathering and analysis techniques. For *data gathering*, out of 44 papers mentioning the techniques used, the most common option (29 papers) was to extract content from a document storage (e.g., documents related to learning/course designs in a learning management system or data from a web page), followed by those using activity tracking and log data (10 papers). In addition, some authors used surveys (5) and interviews (3). Nonetheless, it should be noted that in several papers, only some of the data gathering techniques were mentioned.

In terms of *data* used in the analysis, 25 papers (52.1%) used academic information from the students, 19 (39.6%) learning or course designs, 6 (12.5%) used content downloaded from non-academic websites (e.g., job requirements and forum data), 4 (8.3%) activity traces, 5 (10.4%) other personal data, and 4 (8.3%) relied on learning content generated by the students. Only 2 (4.2%) papers did not collect any data.

Finally, Figure 5 provides an overview of the *analysis techniques* used in the reviewed papers. As we can see, while there is a wide variety of techniques, text mining and descriptive statistics are the most prominent ones.

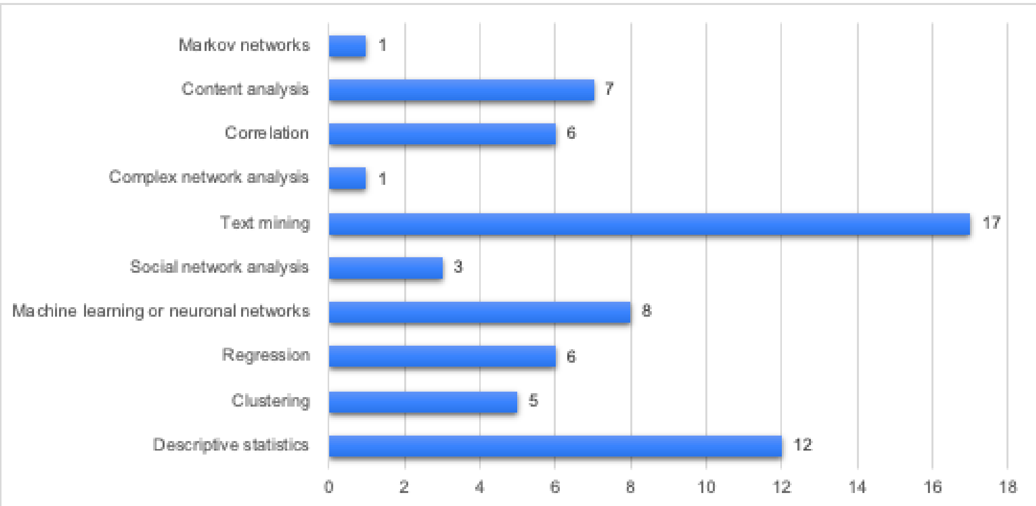


Figure 5: algorithms or techniques used to analyze data

RQ3) What is the level of the maturity of those solutions?

Out of 48 studies only 20 have conducted evaluation. Regarding user involvement, only 8 papers were evaluated with stakeholders, which included students (4 paper), teachers (1 paper), curriculum designers (1 paper), career counsellor (1 paper), and program curators (1 paper). In terms of the type of evaluation, 13 papers evaluated their contributions with already existing data from a real setting, 3 in authentic settings, 1 with a proof of concept, 1 with focus groups, and 1 with experts. For one of the papers, the kind of evaluations was not stated. Finally, regarding the purpose of the evaluation, 12 papers focused on the accuracy, 2 on the usability, 2 on the effectiveness, 1 on the feasibility, 1 on the adoption, and 1 on the performance of the solution.

5 DISCUSSION, CONCLUSIONS AND FUTURE WORK

According to the international and European reports, student success and dropouts constitute a significant concern. Many HEIs are trying to improve teaching practices and the student learning process to address dropouts. However, along with improving teaching strategies, it is necessary to improve the curriculum as well (Gottipati & Shankararaman, 2018) since continuous curriculum improvement provides better results for students and higher education programs (Pistilli and Heileman, 2017). To support that need, this review signals first insights to improve the curriculum through analytics, extending the current works (Ifenthaler & Yau, 2020) by putting more emphasis on the curriculum analytics aspects.

Coming back to the research questions addressed in this paper, the results show the variety of existing *CA solutions proposed in the literature* for HE settings, including theoretical proposals (e.g., such as processes, models, methods, frameworks, and architectures) and practical ones (i.e., tools). However, when we look at how these solutions were used in relation to the curriculum, most of them aimed at understanding it, and just a couple of papers reached the level of optimizing it. Furthermore, the maturity level of those solutions is, in most of the cases, in a very early stage. In fact, only 16.67% of the papers were evaluated with stakeholders and only 6.25% reported evaluations taking place in authentic settings. Thus, further work needs to be done until the adoption of those solutions.

While the presented results do not come without limitations (e.g., due to the query, the selection of databases, bias and inaccuracy in data extraction as it was performed only by one author, or lack of information reported in the papers), based on these results and in connection to the related literature, this paper proposes the following guidelines for the future CA agenda:

- *Theoretical grounding.* In line with the synergies between learning design and analytics, it is important to emphasize that there should be a theoretical ground behind the CA solutions that help stakeholders in the decision making (Macfadyen et al., 2020).
- *Wider variety of CA studies:* At the moment, most of the CA studies focus on analyzing program structure, such as providing the best program path to follow for the desired job or finding the best curriculum path for successful graduation. Further, most of those studies are limited to processes. Very few studies focus on analyzing the curriculum in reflecting faculty teaching and student learning. The available studies are linked to individual students and actions, such as reflecting on their own core competencies corresponding to the covered curriculum. Thus, there is a need for CA tools to understand and improve also other curriculum aspects (e.g., competence-based curriculum assessments).
- *Increase stakeholder involvement*: While Ochoa (2016) presented CA as a solution addressing mainly program managers and directors, in this review we have seen that, while not extensively, other stakeholders such as students, teachers and administrators were taken into account. Still, in order to promote adoption, it would be necessary to further engage the different stakeholders by the CA solutions during the design, deployment and assessment of

the proposed solutions (Rodríguez-Triana et al., 2018; Tsai et al., 2018). This would help to better satisfy the stakeholders needs and to adjust the solutions to their practice.

- *Benefit from visualisations:* Even though most LA studies relate to the development of visualisations (Gašević et al., 2017; Wise et al., 2014), the selected CA papers lack it. In addition, we found that many reviewed studies provide solutions without incorporating them into learning environments, such as learning or institutional management systems. To cover this gap, visualizations could play a helpful role to introduce analytics in help to integrate into different learning environments for when improving curriculum improvement. More concretely, dashboards are one possible solution to provide institutional stakeholders with a real-time picture of the situation (Schwendimann et al., 2016).
- *Benefit from multimodal analytics:* Compared to the other LA reviews (e.g., Ifenthaler & Yau, 2020; Romero & Ventura, 2020), the data sources, data gathering techniques and data analysis techniques are limited in variety. Also, the number of studies including different data sources is scarce. For example, combining stakeholders' feedback, teacher data (observation data, teacher traces), student behavioural data, and course metadata could help to get a broad understanding of the current teaching and learning practices. This points out that the MMLA field may be of great help in order to understand multiple factors conditioning the curriculum.
- *Move from understanding to optimizing.* Most of the CA solutions identified in this review focused on understanding. To move one step forward towards the optimization, if we want to facilitate informed-decision making about the curriculum (Hilliger et al., 2020), it would be necessary to increase the actionability of the CA solutions, e.g., prompting and supporting the interpretation and reflection on the data, and explicitly connecting the retrieved evidence with the decisions that the targeted stakeholders have to make. Also, most of the tools are still in the prototyping phase or implemented on a very small scale. Furthermore, a clear relationship between program outcomes improvement has not been established. In other words, there is still limited research on how program curators accept, interpret and use CA to improve the program outcomes.
- *Further evaluation.* While CA's ultimate goal is to improve student success and reduce dropouts (Mendez et al., 2014), there is still little evidence on that regard. To address this gap, there is a need for more thorough evaluations, including authentic settings and longitudinal studies that show the impact of the solutions in practice. Also, HE institutions would highly benefit from studies that report on the CA solutions from different perspectives (e.g., such as performance, effectiveness, accuracy and usefulness), enabling also comparative studies. For that goal, it would be necessary to define a common framework for CA evaluation.

ACKNOWLEDGMENT

This work was partially funded by the Estonian Research Council (PUT grant PSG286) and the European Union in the context of CEITER (Horizon 2020 Research and Innovation Programme, grant agreement no. 669074).

REFERENCES

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).

Fraser, S. P., & Bosanquet, A. M. (2006). The curriculum? That's just a unit outline, isn't it?. *Studies in higher education*, *31*(03), 269-284.

Gašević, D., Mirriahi, N., Dawson, S., & Joksimović, S. (2017). Effects of instructional conditions and experience on the adoption of a learning tool. *Computers in Human Behavior*, *67*, 207-220.

Gottipati, S., & Shankararaman, V. (2018). Competency analytics tool: Analyzing curriculum using course competencies. *Education and Information Technologies*, *23*(1), 41-60.

Greer, J., Molinaro, M., Ochoa, X., & McKay, T. (2016, April). Learning analytics for curriculum and program quality improvement (PCLA 2016). In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 494-495).

Hernández-Leo, D., Martinez-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.

Hilliger, I., Aguirre, C., Miranda, C., Celis, S., & Pérez-Sanagustín, M. (2020, March). Design of a curriculum analytics tool to support continuous improvement processes in higher education. *In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 181-186).

Ifenthaler, D., & Yau, J. Y. K. (2020). Utilising learning analytics to support study success in higher education: a systematic review. *Educational Technology Research and Development*, *68*(4), 1961-1990.

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Staffordshire, Durham: EBSE 2007-001 Technical Report. Keele University and Durham University Joint Report.

Larrabee Sønderlund, A., Hughes, E., & Smith, J. (2019). The efficacy of learning analytics interventions in higher education: A systematic review. *British Journal of Educational Technology*, 50(5), 2594-2618.

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist*, 57(10), 1439-1459.

Macfadyen, L. P., Lockyer, L., & Rienties, B. (2020). Learning design and learning analytics: Snapshot 2020. *Journal of Learning Analytics*, 7(3), 6-12.

Mangaroska, K., & Giannakos, M. (2018). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. *IEEE Transactions on Learning Technologies*, *12*(4), 516-534.

Mendez, G., Ochoa, X., Chiluiza, K., & De Wever, B. (2014). Curricular design analysis: a data-driven perspective. *Journal of Learning Analytics*, 1(3), 84-119.

Ochoa, X. (2016, April). Simple metrics for curricular analytics. In *CEUR Workshop Proceedings* (Vol. 1590, pp. 20-26). CEUR-WS.

Pistilli, M. D., & Heileman, G. L. (2017). Guiding early and often: Using curricular and learning analytics to shape teaching, learning, and student success in gateway courses. *New Directions for Higher Education*, *2017*(180), 21-30.

Rodríguez-Triana, M. J., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2015). Scripting and monitoring meet each other: Aligning learning analytics and learning design to support teachers in orchestrating CSCL situations. *British Journal of Educational Technology*, *46*(2), 330-343.

Rodríguez-Triana, M. J., Prieto, L. P., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2018, March). The teacher in the loop: Customizing multimodal learning analytics for blended learning. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 417-426).

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), e1355.

Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., ... & Dillenbourg, P. (2016). Perceiving learning at a glance: A systematic literature review of learning dashboard research. IEEE Transactions on Learning Technologies, 10(1), 30-41.

Sclater, N., Peasgood, A., & Mullan, J. (2016). Learning analytics in higher education: A review of UK and international practice. Bristol: JISC

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.

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.

Tinto, V. (2005). Reflections on student retention and persistence: Moving to a theory of institutional action on behalf of student success. *Studies in Learning, Evaluation, Innovation and Developmental Psy-chology*, 2(3), 89–97.

Wise, A. F., Zhao, Y., & Hausknecht, S. N. (2014). Learning analytics for online discussions: Embedded and extracted approaches. *Journal of Learning Analytics*, *1*(2), 48-71.

APPENDIX A - LIST OF REVIEWED PAPERS

Aljohani, N. R., Daud, A., Abbasi, R. A., Alowibdi, J. S., Basheri, M., & Aslam, M. A. (2019). An integrated framework for course adapted student learning analytics dashboard. *Computers in Human Behavior*, 92, 679-690.

Barb, A. S., & Kilicay-Ergin, N. (2020). Applications of Natural Language Techniques to Enhance Curricular Coherence. *Procedia Computer Science*, 168, 88-96.

Bozkurt, I. (2014, June). Quantitative analysis of graduate-level engineering management programs. *In 2014 IEEE International Technology Management Conference* (pp. 1-8). IEEE.

Chang, J. K., Tsao, N. L., Kuo, C. H., & Hsu, H. H. (2016, July). Curriculum Design--A Data-Driven Approach. *In 2016 5th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)* (pp. 492-496). IEEE.

Chen, G., & Yang, J. (2012). The Analysis of University Curriculum Setting Based on KNIME. *IERI Procedia*, 2, 765-770.

Chou, C. Y., Tseng, S. F., Chih, W. C., Chen, Z. H., Chao, P. Y., Lai, K. R., ... & Lin, Y. L. (2015). Open student models of core competencies at the curriculum level: Using learning analytics for student reflection. *IEEE Transactions on Emerging Topics in Computing*, 5(1), 32-44.

Chou, C. Y., Tseng, S. F., Wang, C. J., Chao, P. Y., Chen, Z. H., Lai, K. R., ... & Yu, L. C. (2018). Learning analytics on graduates' academic records to reflect on a competency-based curriculum. *Computer Applications in Engineering Education*, 26(6), 2168-2182.

Constantinov, C., Popescu, P. Ş., Poteraş, C. M., & Mocanu, M. L. (2015, October). Preliminary results of a curriculum adjuster based on professional network analysis. *In 2015 19th International Conference on System Theory, Control and Computing (ICSTCC)* (pp. 860-865). IEEE.

Danubianu, M. (2016, May). First steps towards curriculum development decisions from raw educational data. In 2016 International Conference on Development and Application Systems (DAS) (pp. 281-286). IEEE.

Dobozy, E. (2011). Structured dialogue design in LAMS through interactive lecture podcasting. *Teaching English with technology*, 11(1), 1-18.

Dzemyda, G. (2005). Multidimensional data visualization in the statistical analysis of curricula. *Computational statistics & data analysis*, 49(1), 265-281.

Fortino, A., Zhong, Q., Huang, W. C., & Lowrance, R. (2019, March). Application of Text Data Mining To STEM Curriculum Selection and Development. I*n 2019 IEEE Integrated STEM Education Conference (ISEC)* (pp. 354-361). IEEE.

Gottin, V., Jiménez, H., Finamore, A. C., Casanova, M. A., Furtado, A. L., & Nunes, B. P. (2017, July). An analysis of degree curricula through mining student records. *In 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)* (pp. 276-280). IEEE.

Gottipati, S., Shankararaman, V., & Gan, S. (2017, October). A conceptual framework for analyzing students' feedback. *In 2017 IEEE Frontiers in Education Conference* (FIE) (pp. 1-8). IEEE.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, 15(3), 42-57.

Hilliger, I., Aguirre, C., Miranda, C., Celis, S., & Pérez-Sanagustín, M. (2020, March). Design of a curriculum analytics tool to support continuous improvement processes in higher education.

In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge (pp. 181-186).

Hsia, T. C., Shie, A. J., & Chen, L. C. (2008). Course planning of extension education to meet market demand by using data mining techniques–an example of Chinkuo technology university in Taiwan. Expert Systems with Applications, 34(1), 596-602.

Joshi, A., Desai, P., & Tewari, P. (2020). Learning Analytics framework for measuring students' performance and teachers' involvement through problem based learning in engineering education. *Procedia Computer Science*, 172, 954-959.

Kanuru, S. L., & Priyaadharshini, M. (2020). Lifelong Learning in higher education using Learning Analytics. *Procedia Computer Science*, 172, 848-852.

Kawintiranon, K., Vateekul, P., Suchato, A., & Punyabukkana, P. (2016, December). Understanding knowledge areas in curriculum through text mining from course materials. *In 2016 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)* (pp. 161-168). IEEE.

Knauf, R., Yamamoto, Y., & Sakurai, Y. (2013, December). Optimizing university curricula through correlation analysis. *In 2013 International Conference on Signal-Image Technology & Internet-Based Systems* (pp. 324-329). IEEE.

Kollom, K., Tammets, K., Scheffel, M., Tsai, Y. S., Jivet, I., Muñoz-Merino, P. J., ... & Ley, T. (2021). A four-country cross-case analysis of academic staff expectations about learning analytics in higher education. *The Internet and Higher Education*, 49, 100788.

Lu, L. S., Guan, S. S., Chang, W. S., & Huang, C. C. (2013). A study on the curriculum structure and industry needs for digital media design education in Taiwan. *Procedia-Social and Behavioral Sciences*, 83, 895-899.

Maldonado, E., & Seehusen, V. (2018). Data mining student choices: A new approach to business curriculum planning. *Journal of Education for Business*, 93(5), 196-203.

Mangmang, G. B., Feliscuzo, L., & Maravillas, E. A. Descriptive Feedback on Interns' Performance using a text mining approach. In 2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP) (pp. 1-4). IEEE.

Matulatan, T., & Resha, M. (2017, September). Deep learning on curriculum study pattern by selective cross join in advising students' study path. *In 2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)* (pp. 1-5). IEEE.

Mendez, G., Ochoa, X., Chiluiza, K., & De Wever, B. (2014). Curricular design analysis: a data-driven perspective. *Journal of Learning Analytics*, 1(3), 84-119.

Naujokaitienė, J., Tamoliūnė, G., Volungevičienė, A., & Duart, J. (2020). Using learning analytics to engage students: Improving teaching practices through informed interactions. *Journal of New Approaches in Educational Research (NAER Journal)*, 9(2), 231-244.

Ognjanovic, I., Gašević, D., & Dawson, S. (2016). Using institutional data to predict student course selections in higher education. *The Internet and Higher Education*, 29, 49-62.

Ojha, A., Pattnaik, U., & Sankar, S. R. (2017, August). Data analytics on placement data in a South Asian University. *In 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)* (pp. 2438-2442). IEEE.

Olmos, M. M., & Corrin, L. (2012). Learning analytics: A case study of the process of design of visualizations.

Previde, P., Graterol, C., Love, M. B., & Yang, H. (2019, October). A Data Mining Approach to Understanding Curriculum-Level Factors That Help Students Persist and Graduate. *In 2019 IEEE Frontiers in Education Conference (FIE)* (pp. 1-9). IEEE.

Prinsloo, P., & Slade, S. (2014). Educational triage in open distance learning: Walking a moral tightrope. International Review of Research in Open and Distributed Learning, 15(4), 306-331.

Priyaadharshini, M., Dakshina, R., & Sandhya, S. (2020). Learning Analytics: Game-based Learning for Programming Course in Higher Education. *Procedia Computer Science*, 172, 468-472.

Priyambada, S. A., Mahendrawathi, E. R., & Yahya, B. N. (2017). Curriculum Assessment of Higher Educational Institution Using Aggregate Profile Clustering. *Procedia Computer Science*, 124, 264-273.

Pumpuang, P., Srivihok, A., Praneetpolgrang, P., & Numprasertchai, S. (2008, February). Using bayesian network for planning course registration model for undergraduate students. *In 2008 2nd IEEE International Conference on Digital Ecosystems and Technologie*s (pp. 492-496). IEEE.

Ramingwong, S., Eiamkanitchat, N., & Ramingwong, L. Curriculum Analysis Based on Cerebral Hemisphere Functions Using Association Rule. *In 2018 10th International Conference on Knowledge and Smart Technology (KST)* (pp. 47-52). IEEE.

Rodríguez-Cerezo, D., Sarasa-Cabezuelo, A., Gómez-Albarrán, M., & Sierra, J. L. (2014). Serious games in tertiary education: A case study concerning the comprehension of basic concepts in computer language implementation courses. *Computers in Human Behavior*, 31, 558-570.

Ross, S. R., Volz, V., Lancaster, M. K., & Divan, A. (2018). A Generalizable Framework for Multi-Scale Auditing of Digital Learning Provision in Higher Education. *Online Learning*, 22(2), 249-270.

Saarela, M., & Kärkkäinen, T. (2015). Analysing student performance using sparse data of core bachelor courses. *Journal of educational data mining*, 7(1).

Slim, A., Heileman, G. L., Kozlick, J., & Abdallah, C. T. (2014, December). Employing markov networks on curriculum graphs to predict student performance. *In 2014 13th International Conference on Machine Learning and Applications* (pp. 415-418). IEEE.

Slim, A., Kozlick, J., Heileman, G. L., & Abdallah, C. T. (2014, July). The complexity of university curricula according to course cruciality. *In 2014 Eighth International Conference on Complex, Intelligent and Software Intensive Systems* (pp. 242-248). IEEE.

Vaithilingam, C. A., Gamboa, R. A., & Lim, S. C. (2019). Empowered pedagogy: Catching up with the future. *Malaysian Journal of Learning and Instruction*, 16(1), 1-22.

van der Stappen, E. (2018, April). Workplace learning analytics in higher engineering education. *In 2018 IEEE Global Engineering Education Conference (EDUCON)* (pp. 15-20). IEEE.

Xun, L. S., Gottipati, S., & Shankararaman, V. (2015, June). Text-mining approach for verifying alignment of information systems curriculum with industry skills. *In 2015 International Conference on Information Technology Based Higher Education and Training (ITHET)* (pp. 1-6). IEEE.

Zhu, G., Xing, W., & Popov, V. (2019). Uncovering the sequential patterns in transformative and non-transformative discourse during collaborative inquiry learning. *The Internet and Higher Education*, 41, 51-61.

Zhu, Q., & Wang, M. (2020). Team-based mobile learning supported by an intelligent system: case study of STEM students. Interactive Learning Environments, 28(5), 543-559.

Predicting Early Dropout: Calibration and Algorithmic Fairness Considerations

Marzieh Karimi-Haghighi, Carlos Castillo, Davinia Hernández-Leo, Veronica Moreno Oliver
Universitat Pompeu Fabra
{marzieh.karimihaghighi, carlos.castillo, davinia.hernandez-leo, veronica.moreno}@upf.edu

ABSTRACT: In this work, the problem of predicting dropout risk in undergraduate studies is addressed from a perspective of algorithmic fairness. We develop a machine learning method to predict the risks of university dropout and underperformance. The objective is to understand if such a system can identify students at risk while avoiding potential discriminatory biases. When modeling both risks, we obtain prediction models with an Area Under the ROC Curve (AUC) of 0.77-0.78 based on the data available at the enrollment time, before the first year of studies starts. This data includes the students' demographics, the high school they attended, and their admission (average) grade. Our models are calibrated: they produce estimated probabilities for each risk, not mere scores. We analyze if this method leads to discriminatory outcomes for some sensitive groups in terms of prediction accuracy (AUC) and error rates (Generalized False Positive Rate, GFPR, or Generalized False Negative Rate, GFNR). The models exhibit some equity in terms of AUC and GFNR along groups. The similar GFNR means a similar probability of failing to detect risk for students who drop out. The disparities in GFPR are addressed through a mitigation process that does not affect the calibration of the model.

Keywords: dropout, machine learning, fairness

1 INTRODUCTION

About 36% of university students in the European Union, 39% in the US, 20% in the Australia and New Zealand, and 52% in Brazil discontinue their studies before graduation (Vossensteyn, 2015; Shapiro, 2017; OECD, 2016). Reducing the rate of dropout and underperformance is crucial as these lead to social and financial losses. In addition, detecting students at risk as early as possible is necessary to improve learning and prevent them from quitting and failing their studies.

Research on actionable indicators that can lead to interventions to reduce dropout has received increased attention in the last decade, especially in the Learning Analytics (LA) field (Siemens, 2013; Viberg, 2018; Sclater, 2017; Leitner, 2017). These indicators can help provide effective prevention strategies and personalized intervention actions (Romero, 2019; Larrabee Sønderlund, 2019). Machine Learning (ML) methods, which identify patterns and associations between input variables and the predicted target (Pal, 2012), have been shown to be effective at this predictive task in many LA studies (Plagge, 2013; Kemper, 2020; Aulck, 2016; Nagy, 2018; Del Bonifro, 2020).

We remark that among students who discontinue their studies, some sub-groups are over-represented, something that needs to be considered when developing ML methods. For example, in the UK elder students at point of entry (over 21 years) are more likely to drop out after the first year compared to younger students who enter university directly from high school (Larrabee Sønderlund, 2019). In the US, graduation rate among ethnic minority university students is lower than among White students (Shapiro, 2017). Disparities in risks have been studied in previous work (Gardner, 2019; Hutt, 2019; Kizilcec, 2020) and are addressed in our work by performing per-group analysis of dropout risk and algorithmic bias mitigation of the risk predictions across different groups.

Our contribution. We observe a high dropout rate (43%) among computer engineering undergraduate students at a university in a multinational country in Europe[1] – in comparison to the average EU university students' dropout rate (36%) (Vossensteyn, 2015). In this work, we predict the risk of university dropout and underperformance in this engineering school. Calibrated ML models, having outputs that can be directly interpreted as probabilities for dropout or underperformance, are created using student's features available at the time of enrolment (before students start their studies). It is notable that dropout can also be due to the lack of some qualitative variables in the engineering field, such as motivation or vocational changes (Salas-Morera, 2019) in addition to the institutional rules. We evaluate our models for accuracy and fairness, as model learning may lead to unfairness for some sensitive groups (Corbett-Davies, 2018; Chouldechova A. a., 2018; Barocas, 2017; Mehrabi, 2019; Zou, 2018). Some of the disparities found are addressed through a mitigation procedure (Pleiss, 2017), which seeks to equalize error rates (generalized false positive rate or generalized false negative rate) across groups while preserving the calibration in each group.

The rest of this paper is organized as follows. Section 2 outlines related work. In Section 3, the dataset used in this study is described. The methodology including the ML models and algorithmic fairness analysis are presented in Section 4. Results are given in Section 5, and a procedure to mitigate algorithmic discrimination is used in Section 6. Finally, conclusions and recommendations are presented in Section 7.

2 RELATED WORK

Machine Learning (ML) methods have been used to predict dropout in higher education. In a paper (Aulck, 2016), the impact of ML on undergraduate student retention is investigated by predicting students dropout (defined as not completing at least one undergraduate degree within 6 calendar years of first enrollment). Using students' demographics and academic transcripts, different ML models result in AUCs between 0.66 and 0.73. In another study (Nagy, 2018), an early university dropout is predicted based on available data at the time of enrollment (personal data and secondary school performance) using several ML models with AUCs from 0.62 to 0.81. Similarly, in a recent study (Del Bonifro, 2020), several ML methods are used to predict the dropout of first-year undergraduate students before the student starts the course or during the first year.

Several studies (Chouldechova A. a., 2018; Corbett-Davies, 2018; Barocas, 2017; Mehrabi, 2019; Zou, 2018), have shown that ML models may lead to discriminatory outcomes for some sensitive groups. There are many different definitions of algorithmic fairness (Narayanan, 2018), some of which are incompatible with one another. It is impossible to satisfy all of them simultaneously except in

[1] Country and university name omitted in this version for double-blind review.

pathological cases (such as a perfect classifier), and in general it is impossible to maximize algorithmic fairness and accuracy at the same time (Berk R. , 2019). Hence, there are necessary trade-offs between different metrics (Kleinberg, 2016). Some studies (Hardt, 2016; Zafar, 2017; Woodworth, 2017) try to mitigate potential algorithmic discrimination by introducing a penalization term for unfairness in an objective function to be optimized. Also, several studies (Zemel, 2013; Kamiran, 2009; Kamishima, 2011) tried to approach statistical parity in which the same probability of receiving a positive-class prediction is considered for different groups.

One of the closest studies to ours (Gardner, 2019), considers algorithmic fairness of predictive models of students dropout in MOOCs in terms of accuracy equity using the Absolute Between-ROC Area (ABROCA) metric. The method to improve algorithmic fairness is slicing analysis, which is also used in another study (Hutt, 2019) to analyze fairness across sociodemographic groups in a predictive ML modeling of on-time college graduation. In comparison, in this study we create calibrated ML models that can predict dropout and underperformance risks solely from information available at the time of enrollment, and that have passed through a bias mitigation procedure to avoid error disparities while keeping calibration.

Calibration means that the output of the classifier is not merely a score, but an estimate of the probability of the (adverse) outcome. When we talk about fairness across two groups, we would like this calibration condition to hold for the cases within each of these groups as well. Due to the importance of calibration in risk assessment tools (Berk R. a., 2018; Dieterich, 2016), some previous work has tried to minimize error disparity across groups while maintaining calibration (Pleiss, 2017). In Pleiss et al.'s work, which is closely related to ours but for a different domain, algorithmic bias in a machine learned risk assessment task (criminal recidivism) is minimized by equalizing generalized false positive rates along different racial backgrounds, finding this equalization to be incompatible with calibration. In contrast, in the work presented on this paper, we try to minimize bias in dropout predictive ML models by equalizing error rates (generalized false positive rate or generalized false negative rate) along some sensitive groups while preserving calibration in each group. Finally, we find that equalization along some groups is not entirely incompatible with calibration.

3 DATASET

The anonymized dataset used in this research have been provided by a university in a multinational country in Europe and consists of 881 computer engineering undergraduate students who first enrolled between 2009 and 2017. From this population, 31 cases who did not enroll for the first trimester, 33 students without admission grade, and 150 students without university grade information (students who first enrolled in 2015 are in this group) were removed and finally 667 cases were remained. Two outcome categories are defined; one is dropout and consists of students who enroll in the first year but do not show up in the second year, the other one is underperformance and consists of students who fail 4 or more of the 12 subjects offered in the first year. Out of 667 cases, 286 students drop out and an additional 62 students underperform.

3.1 Per-Group Analysis

The average (base) risk rates of different groups are shown on Table 1. Foreign students have more risks compared to nationals, and the risk of students with lower admission grades is higher than the risk of students with higher admission grades. Naturally, students who fail more subjects and/or who have to take re-sit exams exhibit more risk than their counterparts. There have been two study programs for the total of 60 credits in the first year; plan A (older) with 10 courses and plan B (newer) including 12 courses. In the newer plan, with the aim of improving learning process, there is a course reorganization so that students can experience their first programming course in the first trimester and as can be seen, this change caused lower dropout and dropout/underperformance compared to the older plan.

Table 1: Per-group risk rates. Groups having 10 percentage points or more of risk compared to their counterparts are marked with an asterisk (*).

Group	Size	Risk of Dropout	Risk of Dropout or Underperformance
Female	9%	41%	54%
Male	91%	43%	52%
Nationals[2]	88%	41%	50%
Foreigners	12%	58% *	69% *
Age ≤ 19 (median age)	55%	44%	55%
Age > 19	45%	41%	48%
High school in same state ("in-State")	76%	44%	53%
High school in another state ("out-of-State")	24%	41%	49%
Public high school	42%	44%	55%
Non-public high school	58%	42%	50%
Avg. admission grade ≤ median	50%	49% *	59% *
Avg. admission grade > median	50%	37%	45%
Exam retake (at least once in first year)	87%	47% *	58% *
No exam retake	13%	13%	13%
Course failure (at least once in first year)	85%	47% *	58% *
No course failure	15%	17%	17%
Plan A (older)	74%	46% *	53%
Plan B (newer)	26%	35%	50%
Passed credits ratio[3] ≤ median	50%	70% *	84% *
Passed credits ratio > median	50%	16%	21%

[2] For the purposes of this work, these are students who were born and are resident in the country.

[3] Number of credits passed over total credits during the first year

4 METHODOLOGY

We consider two predictive tasks: predicting dropout and predicting dropout or underperformance.

4.1 ML-based Models

According to the two ground truths (dropout, and dropout or underperformance), separate ML models are created. The feature set for the models consists of demographics (gender, age, and nationality), high school type and location, and average admission grade. Different ML algorithms: logistic regression, multi-layer perceptron (MLP), and support vector machines (SVM) are used to predict dropout risks. ML models are trained using cases enrolled between 2009 to 2013 (409 cases), then tested on students enrolled in 2014, 2016 and 2017 (258 cases). To mitigate the gender imbalance (only 9% of students are women), we use the SMOTE[4] algorithm (Chawla, 2002). We only apply SMOTE on the training set and keep the original class distributions in the test set to ensure valid results.

4.2 Algorithmic Fairness

Parity in the error rates of different groups ("equalized odds") is a well-established method to mitigate algorithmic discrimination in automatic classification (Hardt, 2016; Zafar, 2017; Woodworth, 2017). At the same time, we want to maintain model calibration (Dieterich, 2016; Berk R. a., 2018), as otherwise the same risk estimate carries different meanings and cannot be interpreted equally for different groups. Hence, a relaxation method (Pleiss, 2017) is used in this paper which seeks to satisfy equalized odds or parity in the error rates while preserving calibration. In most cases, calibration and equalized odds are mutually incompatible goals (Chouldechova A. , 2017; Kleinberg, 2016), so in this method it is sought to minimize only a single error disparity across groups while maintaining calibration probability estimates.

If variable $\boldsymbol{x}$ represents a student's features vector, $\boldsymbol{y}$ indicates whether or not the student drops out, $\boldsymbol{G_1}, \boldsymbol{G_2}$ are the two different groups, and $\boldsymbol{h_1}, \boldsymbol{h_2}$ are binary classifiers which classify samples from $\boldsymbol{G_1}, \boldsymbol{G_2}$ respectively, Generalized False Positive Rate (GFPR) and Generalized False Negative Rate (GFNR) are defined as follows (Pleiss, 2017): the GFPR of classifier $\boldsymbol{h_t}$ for group $\boldsymbol{G_t}$ is $\boldsymbol{c_{fp}}(\boldsymbol{h_t}) = \mathbb{E}_{(\boldsymbol{x},\boldsymbol{y})\sim \boldsymbol{G_t}}[\boldsymbol{h_t}(\boldsymbol{x})|\boldsymbol{y} = \boldsymbol{0}]$. This is the average probability of dropout that the classifier estimates for students who do not drop out. Conversely, the GFNR of classifier $\boldsymbol{h_t}$ is $\boldsymbol{c_{fn}}(\boldsymbol{h_t}) = \mathbb{E}_{(\boldsymbol{x},\boldsymbol{y})\sim \boldsymbol{G_t}}[(\boldsymbol{1} - \boldsymbol{h_t}(\boldsymbol{x}))|\boldsymbol{y} = \boldsymbol{1}]$. So the two classifiers $\boldsymbol{h_1}, \boldsymbol{h_2}$ show probabilistic equalized odds across groups $\boldsymbol{G_1}, \boldsymbol{G_2}$ if $\boldsymbol{c_{fp}}(\boldsymbol{h_1}) = \boldsymbol{c_{fp}}(\boldsymbol{h_2})$ and $\boldsymbol{c_{fn}}(\boldsymbol{h_1}) = \boldsymbol{c_{fn}}(\boldsymbol{h_2})$. Classifier $\boldsymbol{h_t}$ is said to be well-calibrated if $\forall \boldsymbol{p} \in [\boldsymbol{0}, \boldsymbol{1}], \mathbf{P}_{(\boldsymbol{x},\boldsymbol{y})\sim \boldsymbol{G_t}}[\boldsymbol{y} = \boldsymbol{1}|\boldsymbol{h_t}(\boldsymbol{x}) = \boldsymbol{p}] = \boldsymbol{p}$. To prevent the probability scores from carrying group-specific information, both classifiers $\boldsymbol{h_1}, \boldsymbol{h_2}$ are also calibrated with respect to groups $\boldsymbol{G_1}, \boldsymbol{G_2}$ (Berk R. a., 2018; Dieterich, 2016).

[4] Synthetic Minority Oversampling Technique

5 RESULTS

5.1 Effectiveness Evaluation

The best results for both dropout risk predictions were obtained using a Multi-Layer Perceptron (MLP). We used a single hidden layer having 100 neurons. The other models are omitted for brevity. Results in terms of the AUC-ROC, GFNR, GFPR, and F-score (the harmonic mean of precision and recall, which unlike the other metrics, requires to establish an optimal cut-off for classification) are presented in Table 2. According to the results, the models lead to good performance in terms of AUC and F-score in both prediction tasks. With a little information at the time of students' enrollment, these models show good AUC in comparison to previous work (Aulck, 2016; Nagy, 2018) which showed AUC in the order of 0.62-0.81. Also, comparing calibrated and non-calibrated predictions we can see that calibrated model leads to lower GFNR and non-calibrated results in lower GFPR.

5.2 Algorithmic Fairness Evaluation

The results for the analysis of algorithmic fairness are shown on the left side of Table 3. In dropout prediction, we can observe accuracy equity (less than 20% discrepancy) in terms of AUC in both models, even if results are slightly more accurate for male students. AUC is also higher for students with lower admission grades compared to their counterparts. In the calibrated model, males, foreigners, and lower admission grade students experience lower GFNR compared to their counterparts. However, non-calibrated model shows fairer results for GFNR along these groups. Regarding GFPR, there can be seen more false positive errors (higher risk scores for students who do not dropout or underperform) for males compared to females, students of out-of-State high schools than in-State high schools, and lower admission grade students compared to their counterparts in the non-calibrated model. In the calibrated model, this metric shows more errors for foreigners and for lower admission grade students compared to their counterparts.

Similar results are shown for predicting dropout or underperformance. In terms of AUC, MLP shows equity (less than 20% discrepancy) across groups except for more accuracy for students from in-State high schools. In the calibrated model, higher AUC can be observed in nationals compared to foreigners and higher admission grade students. Also, both models show parity across all groups in terms of GFNR except for students with lower admission grade who experience lower errors compared to their counterparts, however, non-calibrated model shows lower discrimination to this groups compared to the calibrated one. In terms of GFPR, we can see more errors of the model for foreigners than nationals, out-of-State high school than in-State high school students, males than females, and cases with lower admission grades compared to their counterparts. In the calibrated model, this metric also shows more error for foreigners than national and students with lower admission grade compared to their counterparts, but it reveals more errors for females than males.

Table 2: Effectiveness of models in risk prediction.

Risk	Dropout				Dropout or Underperformance			
Model	AUC	GFNR	GFPR	F-score	AUC	GFNR	GFPR	F-score
MLP	0.77	0.73	0.19	0.76	0.78	0.69	0.19	0.83
MLP calibrated	0.77	0.36	0.42	0.76	0.78	0.27	0.49	0.83

6 EQUALIZED ODDS AND CALIBRATION

In this section, parity is sought along groups in terms of two fairness metrics. For this purpose, the method introduced by Pleiss et al. (Pleiss, 2017) is used, which seeks parity in Generalized False Positive Rate (GFPR) or Generalized False Negative Rate (GFNR) while preserving calibration. In both prediction tasks, the models before mitigation exhibit in general better parity in terms of AUC and GFNR and more inequality in terms of GFPR. The results after bias mitigation are presented in the right side of the Table 3. By comparing the results before and after GFPR bias mitigation in dropout we can see that the disparity in GFPR has decreased in the order of 0.03-0.71 in MLP and 0.02-0.30 in MLP calibrated across all groups. Also, comparing the result before and after GFPR bias mitigation in dropout or underperformance show that bias in MLP and MLP calibrated models has been respectively reduced by the order of 0.08-1.15 and 0.14-0.59 across all groups.

7 CONCLUSIONS AND RECOMMENDATIONS

The effectiveness and fairness of Machine Learning (ML) models in the early prediction of university dropout and underperformance was evaluated. Using only information at the time of enrollment, calibrated ML models were created with AUC of 0.77 and 0.78 which can help reliably identify students at risk to trigger interventions that can help increase their success and ultimately reduce social and economic costs. When introducing ML models, improvements in accuracy need to be carefully contrasted with potential algorithmic discrimination. Thus, we evaluated the algorithmic fairness of the ML models in terms of AUC and error (GFNR and GFPR) across five groups defined by nationality, gender, high school type and location, and admission grade. According to the results, our modeling has parity in terms of AUC and GFNR but disparities in GFPR. These disparities in GFPR are larger among groups defined by admission grade, and the bias is against students with lower admission grades. The predicted probability of dropout for the students of this sub group who do not actually drop out is larger than that of their counterparts (students of higher admission grade sub group). Using a relaxation method (Pleiss, 2017), we tried to obtain parity in GFPR while preserving calibration. By maintaining the calibration among subgroups, we prevent the probability scores from needing group-dependent interpretation. The results after bias mitigation show that GFPR ratio in both dropout and dropout or underperformance predictions has been changed to a perfect value close to 1 across most of the groups. This bias mitigation also caused better parities in other metrics (AUC and GFNR) along majority of the groups compared to the non-mitigated model. Studying algorithmic discrimination means addressing unfair decisions not only to the identification of students that would require preventive mentoring programs, but also to the identification of potentially successful students that would benefit from e.g. additional educational opportunities or to the formulation of pedagogical interventions related to changes in the study plans or in pedagogical methods suiting specific students' profiles.

In terms of contributions to learning analytics, in addition to creating ML models for dropout and underperformance that exhibit high accuracy, we evaluated algorithmic fairness of the models across different groups in terms of several metrics and applied a bias mitigation method to set parity for subgroups with unfair results. For the students at high risk of dropout or underperformance, different interventions can be considered such as tutoring, counselling and mentoring. A suggested beneficial intervention (Lowis, 2008) is interviewing with the students in informal discussion and asking for their perceptions and experiences at the university which can help with the planning process for their subsequent academic years. Also, a preventive mentoring program (Larose, 2011) showed high levels of motivation and more positive career decision profiles for the newcomer students who participated in bimonthly meetings with students completing their undergraduate degree. Both require early prediction models with equity among groups, which the methods we have described can provide in a real-world setting.

ACKNOWLEDGMENTS

This work has been partially supported by the HUMAINT programme (Human Behavior and Machine Intelligence), Centre for Advanced Studies, Joint Research Centre, European Commission. The project leading to these results have received funding from "la Caixa" Foundation (ID 100010434), under the agreement LCF/PR/PR16/51110009. We also acknowledge the support of ICREA Academia.

REFERENCES

Aulck, L. a. (2016). Predicting student dropout in higher education. *ICML Workshop on #Data4Good: Machine Learning in Social Good Applications*.

Barocas, S. a. (2017). Fairness in machine learning. *NIPS Tutorial*.

Berk, R. (2019). Accuracy and fairness for juvenile justice risk assessments. *Journal of Empirical Legal Studies*, 175-194.

Berk, R. a. (2018). Fairness in criminal justice risk assessments: The state of the art. *Sociological Methods & Research*.

Chawla, N. V. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 321-357.

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, 153-163.

Chouldechova, A. a. (2018). The frontiers of fairness in machine learning. *arXiv preprint arXiv:1810.08810*.

Corbett-Davies, S. a. (2018). The measure and mismeasure of fairness: A critical review of fair machine learning. *arXiv preprint arXiv:1808.00023*.

Del Bonifro, F. a. (2020). Student Dropout Prediction. *International Conference on Artificial Intelligence in Education*, (pp. 129-140).

Dieterich, W. a. (2016). COMPAS risk scales: Demonstrating accuracy equity and predictive parity. *Northpoint Inc*.

Gardner, J. a. (2019). Evaluating the fairness of predictive student models through slicing analysis. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge.*

Hardt, M. a. (2016). Equality of opportunity in supervised learning. *Advances in neural information processing systems*, (pp. 3315-3323).

Hutt, S. a. (2019). Evaluating Fairness and Generalizability in Models Predicting On-Time Graduation from College Applications. *Proceedings of The 12th International Conference on Educational Data Mining*, (pp. 79-88).

Kamiran, F. a. (2009). Classifying without discriminating. *2009 2nd International Conference on Computer, Control and Communication* (pp. 1-6). IEEE.

Kamishima, T. a. (2011). Fairness-aware learning through regularization approach. *2011 11th IEEE International Conference on Data Mining Workshops*, (pp. 643-650).

Kemper, L. a. (2020). Predicting student dropout: A machine learning approach. *European Journal of Higher Education*, 28-47.

Kizilcec, R. F. (2020). Algorithmic Fairness in Education. *Ethics in Artificial Intelligence in Education*.

Kleinberg, J. a. (2016). Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807*.

Larose, S. a. (2011). Academic mentoring and dropout prevention for students in math, science and technology. *Mentoring \& Tutoring: Partnership in Learning*, 419-439.

Larrabee Sønderlund, A. a. (2019). The efficacy of learning analytics interventions in higher education: A systematic review. *British Journal of Educational Technology*, 2594-2618.

Leitner, P. a. (2017). *Learning analytics in higher education—a literature review.* Springer.

Lowis, M. a. (2008). Factors affecting student progression and achievement: prediction and intervention. A two-year study. *Innovations in education and teaching international*.

Mehrabi, N. a. (2019). A survey on bias and fairness in machine learning. *arXiv preprint arXiv:1908.09635*.

Nagy, M. a. (2018). Predicting dropout in higher education based on secondary school performance. *2018 IEEE 22nd international conference on intelligent engineering systems (INES).*

Narayanan, A. (2018). 21 fairness definitions and their politics. *presenterad på konferens om Fairness, Accountability, and Transparency*.

OECD. (2016). *Education at a Glance 2016: OECD Indicators.* Paris: OECD, http://dx.doi.org/10.187/eag-2016-en.

Pal, S. (2012). Mining educational data to reduce dropout rates of engineering students. *International Journal of Information Engineering and Electronic Business*.

Plagge, M. (2013). Using artificial neural networks to predict first-year traditional students second year retention rates. *Proceedings of the 51st ACM Southeast Conference*, (pp. 1-5).

Pleiss, G. a. (2017). On fairness and calibration. *Advances in Neural Information Processing*.

Romero, C. a. (2019). Guest editorial: special issue on early prediction and supporting of learning performance. *IEEE Transactions on Learning Technologies*, 145-147.

Salas-Morera, L. a.-H.-R. (2019). Factors affecting engineering students dropout: a case study. *The International journal of engineering education*, 156-167.

Sclater, N. a. (2017). Learning analytics in higher education. *London: Jisc. Accessed February*, 176.

Shapiro, D. a. (2017). *Completing college: A national view of student completion rates--fall 2011 cohort.* National Student Clearinghouse Research Center.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*.

Viberg, O. a. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior*, 98-110.

Vossensteyn, J. J. (2015). *Dropout and completion in higher education in Europe: Main report.* European Union.

Woodworth, B. a. (2017). Learning non-discriminatory predictors. *Proceedings of Machine Learning Research*.

Zafar, M. B. (2017). Fairness beyond disparate treatment \& disparate impact: Learning classification without disparate mistreatment. *Proceedings of the 26th international conference on world wide web*, (pp. 1171-1180).

Zemel, R. a. (2013). Learning fair representations. *International Conference on Machine Learning.*

Zou, J. a. (2018). AI can be sexist and racist—it's time to make it fair. *Nature Publishing Group*.

Table 3: Effectiveness (AUC) and fairness (GFPR and GFNR ratios) of models for the two risk prediction tasks, before and after bias mitigation. Values in boldface should, ideally, be close to 1.0 to indicate perfect equity among groups.

	Before bias mitigation												After bias mitigation											
Risk	Dropout						Dropout or Underperformance						Dropout						Dropout or Underperformance					
Model	MLP			MLP calibrated			MLP			MLP calibrated			Equalized GFPR MLP			Equalized GFPR MLP calibrated			Equalized GFPR MLP			Equalized GFPR MLP calibrated		
Group / Metric	AUC	GFNR	GFPR	AUC	GFNR	GFPR	AUC	GFNR	GFPR	AUC	GFNR	GFPR	AUC	GFNR	GFPR	AUC	GFNR	GFPR	AUC	GFNR	GFPR	AUC	GFNR	GFPR
Nationals	0.77	0.74	0.19	0.76	0.34	0.42	0.77	0.72	0.19	0.81	0.23	0.45	0.77	0.74	0.19	0.50	0.46	0.54	0.64	0.68	0.28	0.50	0.35	0.65
Foreigners	0.82	0.73	0.17	0.69	0.26	0.61	0.67	0.66	0.25	0.55	0.21	0.67	0.82	0.72	0.17	0.69	0.26	0.61	0.67	0.66	0.25	0.55	0.21	0.67
$\frac{\text{Nationals}}{\text{Foreigners}}$ (Ratio)	**0.93**	**1.01**	**1.13**	**1.10**	**1.33**	**0.68**	**1.15**	**1.08**	**0.76**	**1.45**	**1.07**	**0.67**	**0.93**	**1.03**	**1.13**	**0.73**	**1.79**	**0.88**	**0.95**	**1.03**	**1.14**	**0.90**	**1.63**	**0.98**
State_Highschool	0.79	0.74	0.18	0.74	0.35	0.46	0.79	0.72	0.17	0.78	0.22	0.47	0.65	0.71	0.25	0.74	0.35	0.46	0.62	0.62	0.32	0.72	0.27	0.54
NonState_Highschool	0.71	0.70	0.23	0.77	0.31	0.43	0.59	0.67	0.30	0.74	0.25	0.54	0.71	0.70	0.23	0.78	0.30	0.43	0.59	0.67	0.30	0.74	0.25	0.54
$\frac{\text{State\_Highschool}}{\text{NonState\_Highschool}}$ (Ratio)	**1.11**	**1.06**	**0.77**	**0.96**	**1.12**	**1.07**	**1.35**	**1.08**	**0.57**	**1.05**	**0.88**	**0.87**	**0.92**	**1.02**	**1.07**	**0.94**	**1.13**	**1.09**	**1.06**	**0.93**	**1.08**	**0.97**	**1.06**	**0.99**
Pub_Highschool	0.82	0.74	0.17	0.77	0.36	0.39	0.79	0.71	0.18	0.78	0.22	0.44	0.80	0.71	0.20	0.63	0.41	0.50	0.81	0.68	0.18	0.75	0.24	0.51
NonPub_Highschool	0.72	0.73	0.21	0.72	0.31	0.47	0.72	0.70	0.21	0.75	0.20	0.49	0.72	0.73	0.21	0.72	0.31	0.47	0.72	0.70	0.21	0.75	0.20	0.49
$\frac{\text{Pub\_Highschool}}{\text{NonPub\_Highschool}}$ (Ratio)	**1.14**	**1.01**	**0.85**	**1.07**	**1.19**	**0.82**	**1.09**	**1.01**	**0.88**	**1.03**	**1.07**	**0.91**	**1.10**	**0.96**	**0.96**	**0.88**	**1.33**	**1.05**	**1.11**	**0.97**	**0.88**	**0.99**	**1.18**	**1.05**
Low_AdmissionGrade	0.69	0.70	0.27	0.67	0.24	0.58	0.66	0.65	0.33	0.55	0.15	0.77	0.69	0.70	0.27	0.67	0.24	0.58	0.66	0.65	0.33	0.55	0.15	0.77
High_AdmissionGrade	0.52	0.84	0.16	0.71	0.50	0.37	0.64	0.82	0.15	0.74	0.41	0.40	0.52	0.73	0.27	0.50	0.57	0.43	0.49	0.67	0.34	0.50	0.51	0.49
$\frac{\text{Low\_AdmissionGrade}}{\text{High\_AdmissionGrade}}$ (Ratio)	**1.32**	**0.83**	**1.68**	**0.95**	**0.49**	**1.59**	**1.03**	**0.79**	**2.16**	**0.74**	**0.37**	**1.95**	**1.32**	**0.96**	**1.01**	**1.35**	**0.42**	**1.36**	**1.33**	**0.97**	**0.97**	**1.10**	**0.30**	**1.57**
Male	0.78	0.73	0.20	0.75	0.33	0.45	0.77	0.70	0.20	0.78	0.23	0.47	0.78	0.73	0.20	0.75	0.33	0.45	0.77	0.70	0.20	0.51	0.34	0.65
Female	0.62	0.84	0.13	0.57	0.51	0.44	0.67	0.80	0.15	0.71	0.23	0.64	0.62	0.76	0.20	0.67	0.55	0.45	0.56	0.80	0.22	0.71	0.23	0.64
$\frac{\text{Male}}{\text{Female}}$ (Ratio)	**1.26**	**0.87**	**1.51**	**1.33**	**0.66**	**1.02**	**1.15**	**0.87**	**1.37**	**1.10**	**0.99**	**0.73**	**1.26**	**0.96**	**1.00**	**1.13**	**0.61**	**1.01**	**1.38**	**0.87**	**0.91**	**0.72**	**1.45**	**1.02**

Eliciting Students' Needs and Concerns about a Novel Course Enrollment Support System

Kerstin Wagner[1], Isabel Hilliger[2], Agathe Merceron [1], Petra Sauer [1]
Beuth University of Applied Sciences Berlin, Germany[1]
Pontificia Universidad Católica de Chile, Santiago, Chile[2]
kerstin.wagner@beuth-hochschule.de, ihillige@ing.puc.cl,
merceron@beuth-hochschule.de, sauer@beuth-hochschule.de

ABSTRACT: Selecting courses that optimally fit a student's situation can help reduce the risk of dropping out. Data exploration and performance prediction approaches can be applied to help students make these decisions. To ensure that an enrollment support system meets the needs of students, they should be involved as early as possible in the development process. This paper presents an initial assessment of some functionalities of a novel course enrollment support system based on student performance data. The results include a collection of indicators and sources of information, as well as an overview of needs and concerns. The insights gathered will help to develop a system that has the trust of students.

Keywords: Learning Analytics, Student-Centered Design, Course Enrollment Support

1 INTRODUCTION

One of the most promising uses of learning analytics is the personalization of students' learning experiences. In these lines, recommender systems are used to provide students with different courses of action (Wilson et al., 2017). Recommender systems here refer to the use of computational techniques to make suggestions when there is a great volume of options, as in such situations the selection process can become difficult for the user (Goncalves et al., 2018). In educational settings, recommender systems have been proposed to assist students in choosing courses and learning materials, aiming at improving students' persistence and achievement (Abdi et al., 2020; Goncalves et al., 2018).

Why thinking about a novel course enrollment support system? When students have to choose courses at the beginning of the semester, current practice in many German universities is as follows: first, students look up in their transcripts the courses that they have already completed; second, they consult the university course catalogue and the schedule to see which courses are offered, and when they decide how many courses they would like to follow, they choose and enroll. Presently, there is seldom specific official support put in place by the universities to provide more information or to help students reflect on their studies. Experience shows that, before enrolling, many students look for additional information or advice by talking to friends or fellow students. Very few talk to an advisor.

More support could be provided using student-facing learning analytics to ponder on questions like: How many courses should I take? When is the best time to repeat a failed course? Which grades can I expect? Which courses might be difficult for me? Am I at risk of dropping out? Many students are not aware of the advances in learning analytics and the possibilities recommender systems. Further studies should explore current practices and needs, discussing new possibilities when planning the design of a novel enrollment system that could provide such support, helping to reduce the number of dropouts.

Considering that recent work has highlighted the importance of designing transparent recommender systems to improve learners' satisfaction and trust (Abdi et al. 2020), this paper presents the methodology and the results of the testing of the first loop of a student-centered design approach to implement a course enrollment support system with a recommendation component.

2 RELATED WORK

The authors of several studies have described their user-centered approach to involve students in the development of learning analytics tools. The importance of this was also emphasized, e.g., to make learning analytics tools comparable and to evaluate their impact on students.

Bodily and Verbert (2017) conduct a literature review with 94 articles on systems that track learning analytics data and provide their output directly to students. Considering the functionality, the systems were primarily for awareness or reflection (37%), for resources recommendations (29%), for improving retention or engagement (19%), and a few for course recommendations (3%). The authors noticed that the recommendation components are in many cases not transparent in the way they operate, which harms trust and acceptance. They indicated that, although 37% of systems offer grade comparisons, the question of interpretation is open: students with above-average performance might be careless and below-average frustrated. Furthermore, it was mentioned that despite being systems for students, only 6% of the articles included students as stakeholders in the requirements analysis. Finally, Bodily and Verbert (2017) provide 10 questions to guide the development of a possible system and to create awareness that these issues should be published more extensively and transparently in research papers to enable conclusions about student success.

Jivet et al. (2018) summarized the findings of their systematic literature review of learner dashboard development including 26 papers and gave recommendations for dashboard design and evaluation. For example, one should consider that comparisons with fellow students do not necessarily have a positive impact and no learner group should benefit more than others. In addition to usefulness and ease of use, the authors stated that understanding the data and how to interpret it, and finally, trust in the tool plays a major role in dashboard evaluation.

De Quincey et al. (2019) included students in the development of a dashboard that integrates study motivation to track engagement and predicted scores at Keele University (UK). For iterating through analysis, design, and development, four student ambassadors were trained as user experience researchers, who in turn recruited students to elicit feedback. The initial implementation of the system was tested with 94 volunteer students and then evaluated in 10 contextual interviews. Although trust in the system was rated differently, most students recognized the support provided by the tool and for some it also influenced engagement. The authors list a set of success factors such as integrating the system into the classroom, trust in data and computation, and personalization since not all students share the same goals.

Hilliger et al. (2020) identified student information needs regarding course enrollment at Pontifical Catholic University (Chile) using a mixed-methods approach in which a qualitative survey with open-ended questions and 31 student representatives as participants was turned into a quantitative, closed-ended survey with 627 participants in the second step. Information needs were divided into used information and information sources, and additional required information. The most frequently used sources of information out of 9 are class scheduling dashboards (98%) and visualizations of progress in the program, and the least frequently used are other people (12%), i.e. advisors and acquaintances, and friends. The two most relevant pieces of information out of 12 were course schedules (95%) and program progress (66%) and the least relevant are pending academic credits and course assessment tools. Despite the variety of information and sources already available, there were still items that can support course enrollment decisions, such as the information about the real use of teaching assistants' hours (61%), assessment tool types (52%), and course grades in previous semesters (49%). (Hilliger et al., 2020) recognized desired course-level indicators as mostly descriptive and refered to the rejection of predictive indicators mentioned in former papers.

Sarmiento et al. (2020) described their approach of a series of co-design workshops for learning analytics tools with and for students at New York University (USA). Based on in-depth interviews, personas were developed to map the biggest challenges of the learning experiences. Students were invited to co-design workshops to develop solutions for these personas. Students' time constraints at the end of the semester were not only a problem for recruitment, but also for participation in three 5-hour workshops. Out of 106 students initially considered as potential participants, 20 ended up being interested and finally, a total of 10 participants joined the workshops with 4-7 participants each.

3 METHODOLOGY

As the first test phase with users after the design and development of low-fidelity prototype functionalities (Ladner, 2015), a semi-structured group discussion (SSGD) was used to involve students in the development process. The SSGD was organized in two parts: a general part 1 about needed information and a more specific part 2 regarding descriptive and predictive analytics based on real data. The questions of part 2 were extracted from previous research: the feasibility analysis of a course recommendation system based on grade predictions (Wagner et al., 2020a) and the analysis of students at risk of dropping out (Wagner et al., 2020b).

3.1 Group Discussion

The SSGD was held as part of a seminar in an elective machine learning course planned in the 4th/5th semester of the curriculum. In this course, the 25 students had already worked in seven groups on a variety of topics and had built some understanding in data exploration and statistical learning. Thus, their data literacy level may be considered above average. Part 2 provides students an opportunity to think about the impact of machine learning algorithms on users, which fits well in such a course. Further, by integrating the discussion into the course and awarding points for participation, student time constraints as in (Sarmiento et al., 2020) were overcome.

The two parts of the SSGD were presented and the students discussed in groups for 30 minutes. All seven groups worked on part 1, which was similar to the 1st survey of Hilliger et al. (2020): 1. "What information do you typically use to decide which courses to take?" and 2. "What additional information would you like to have and why?" Part 2 was divided into three focus topics that were not completely free of overlap regarding their tasks: A Descriptive Statistics, B Grade Predictions, and C Dropout Predictions. Each topic was divided into four tasks, consisting of several questions supported by figures. Table 1 gives an overview of topics and tasks and the corresponding figures. The figures shown below are examples based on student 2, who had a high dropout risk in contrast to the others, since not all figures can be shown. The slides as provided to the students can be found online.[1]

Table 1: Group Discussion Questions: Topics from A to C and tasks from A1 to C4

A Descriptive Statistics	**B Grade Predictions**	**C Dropout Predictions.**
A1 Performance Do you find a comparison of your own performance with that achieved by other students helpful, for example, to better assess your own performance? (similar to Figure 1b)	**B1 Information Demand** Do you think that there are different information needs depending on the performance, i.e. does student 2 need different information than student 1? (similar to Figure 1a)	**C1 Individual Study Conditions** Do you think that a recommendation system should be able to take different study motivations / framework conditions into account? Which parameters can you think of / do you find important? (similar to Figure 1a)
A2 Performance Do you find explanatory texts helpful, e.g. to better assess your own performance? (similar to Figure 1b)	**B2 Grades Distribution** Do you find a rough overview of the results in the modules of the coming semester helpful? What evaluations would you be interested in? (similar to Figure 3)	**C2 Dropout Risk** Especially at the beginning of their studies, students often change their study program or drop out. Do you find a corresponding forecast helpful? Should explanations be provided as to how the system makes the forecast? What kind of explanations could be helpful for students to better manage their studies? (Figure 2b)
A3 Grades Distribution Do you find a rough overview of the results in the modules of the coming semester helpful? (Figure 3)	**B3 Grade Predictions** Do you find a grade prediction for the upcoming semester helpful? What effect might the prediction of a good grade, a bad grade, or the prediction of not passing have on you? Should only the positive predictions be given? (similar to Figure 2a)	**C3 Support in Case of Dropout Risk** What support would you want in such a situation?
A4 Further Exploration What other evaluations would interest you?	**B4 Grade Predictions** Do you need explanations to understand the prediction? If so, which ones? If the explanations do not seem helpful to you, would you need a simulator to enter your desired grade, and understand the effect on your degree? (similar to Figure 2a)	**C4 Study Behavior Illustration** Do you find examples of students in similar situations helpful? How could this be illustrated? (similar to Figure 1c)

Two groups focused on topic A, three on topic B and two on topic C. The opinions and ideas were then presented by one group at a time in a whole class discussion and supplemented by the other groups or individual students as needed. The groups provided their notes, which serve as the basis for this paper. All groups were given the enrollments and grades of the first two semesters of three student-examples drawn from the data of their own study program, so that they could easily understand the three examples. Student 1 had almost only very good grades, while student 2 enrolled in fewer

[1] Further material: https://projekt.beuth-hochschule.de/index.php?id=4863

courses and failed one course in semester 1 and semester 2. Student 3 enrolled in the same courses as student 1 and got mostly good grades.

Figure 1a shows the enrollments and grades for student 2. The column "S" means the study semester of a given student, the column "P" is the semester in which a course is planned in the curriculum, the column "Course" is the code of a course, with "WP" representing an elective course and "B" a mandatory course, and the column "Grade" is the grade obtained by the student. Grades are given according to the German system: 1.0 is the best (green), 4.0 the worst possible grade to pass a course (orange) and 5.0 means fail (red). If a course has been enrolled but no exam has been taken, the cell is gray. The last row means: he/she completed the course WP01 in his/her 1st semester of study with the grade of 1.3, and this course is planned in the 4th or 5th semester according to the curriculum. Figure 1b compares the student's results with the course median based on all students and all semesters and gives the difference (red less, green better, gray equal to the course median). Figure 1c shows the results of the remaining semesters.

a)

S	P	Course	Grade
1	1	B01	2.3
1	1	B02	2.7
1	1	B03	1.7
1	1	B04	2.0
1	1	B05	2.0
2	2	B06	1.3
2	2	B07	4.0
2	2	B08	2.7
2	2	B09	2.0
2	2	B10	2.0
2	4/5	WP01	1.7

b)

S	P	Course	Grade	Median	Diff
1	1	B01	3.3	2.3	1.0
1	1	B02	2.7	2.7	0.0
1	1	B03	3.0	1.7	1.3
1	1	B04	2.0	2.0	0.0
1	1	B05	5.0	2.3	2.7
2	1	B05	2.7	2.3	0.4
2	2	B06	5.0	2.7	2.3
2	2	B10	1.7	2.0	-0.3
2	4/5	WP01	1.7	1.7	0.0

c)

S	P	Course	Grade
3	2	B07	4.0
3	2	B09	
3	3	B15	2.0
3	3	B17	2.7
3	4/5	WP06	1.7
4	2	B06	3.3
4	2	B08	1.7
4	2	B09	1.3
4	3	B13	3.7
4	4	B20	2.0
4	5	B24	1.0

S	P	Course	Grade
6	3	B14	3.0
6	3	B16	3.7
6	4	B18	2.7
6	4	B19	1.7
6	5	B23	2.0
6	4/5	WP14	
6	4/5	WP15	
7	4/5	WP03	
7	4/5	WP10	1.7
8	4/5	WP09	3.3

Figure 1: Presentation of the Students' Results using the Example of Student 2:
a) Results from the first two semesters (used in tasks B1, B3, B4, C1, C2, C4),
b) Performance summary of 1st and 2nd student semesters comparing grades of the student to the course median and reports the difference between students grade and course median (used in task A1),
c) Results from later semesters (used in task C4)

Figure 2 combines two prediction results used for topics B and C. The grades prediction result for each course in the upcoming 3rd semester based on linear regression are shown for student 2 in Figure 2a. The grades are colored with the same scheme as in Figure 1. The dropout risk, or probability of graduating, of each student based on logistic regression, are shown in Figure 2b.

a)

Course	Grade
B13	2.33
B14	2.12
B15	2.14
B16	3.09
B17	2.07

b)

Student	Prediction	Probability
1	Graduate	89.07%
2	Dropout	72.35%
3	Graduate	91.30%

Figure 2: Presentation of the Prediction Results:
a) Grade predictions of 3rd semester courses for example student 2 (used in task B3, B4),
b) Dropout prediction for all example students (used in task C2)

Figure 3 shows a summary of the statistics of the students' performance of all courses planned in the 3rd semester: minimum, maximum grades, median, mode, and percentage of the students who got each grade from 1.0 to 4.0 and colored as a heatmap. These statistics have been calculated with historical data.

P	Course	Min - Max	Median	Mode	1.0	1.3	1.7	2.0	2.3	2.7	3.0	3.3	3.7	4.0
3	B13	1.0 - 4.0	2.0	1.3	5.9%	17.8%	17.5%	14.1%	16.6%	10.2%	9.4%	5.2%	1.7%	1.7%
3	B14	1.0 - 4.0	2.0	1.0	25.4%	11.0%	12.5%	12.2%	10.3%	9.6%	5.4%	7.5%	3.7%	2.4%
3	B15	1.0 - 4.0	2.3	2.3	6.8%	5.5%	11.9%	12.9%	16.5%	14.4%	16.4%	12.2%	3.0%	0.5%
3	B16	1.0 - 4.0	2.7	3.0	6.3%	6.9%	7.4%	10.7%	11.8%	9.8%	15.7%	9.0%	9.3%	13.1%
3	B17	1.0 - 4.0	1.7	1.3	10.8%	24.7%	22.3%	16.1%	9.2%	7.1%	4.2%	2.6%	2.1%	1.1%

Figure 3: Summary Statistics of Grades Distributions
for Courses of the 3rd Program Semester(used in task A3):
P = Plan semester (semester in which the course is scheduled according to the curriculum),
Min-Max = Grades range, [1.0, 1.3, …, 4.0] = Grades and their distribution over all semester

3.2 Evaluation

The two parts of group work and discussion have been evaluated differently so far: the submitted notes from part 1 were coded in several steps based on the indicators and information sources from Hilliger et al. (2020), and those from part 2 were used to describe key challenges for further work.

To code the notes from part 1, the most important indicators and sources given by Hilliger et al. (2020) were used as starting point. In the first step, the most comprehensive answer to each question was mapped to the codes by two researchers together. The other notes were coded independently, and those that did not lead to the same result, e.g., because of to unclear meaning due to other study conditions, were discussed afterward. If previous codes could not be adopted, they were merged or renamed. If necessary, new codes were created. The mentioned indicators and sources were evaluated by their number of occurrences and assessed in terms of their integrability into a novel enrollment support system. For part 2, students' comments were grouped into addressed subjects and improvement ideas are briefly outlined.

4 RESULTS

4.1 Part 1

The presentations by the groups and discussions with the students lasted about 15 minutes for part 1. Table 2 gives an overview of the original codes (Hilliger et al., 2020) for indicators as well as the used codes in this paper and their short description and Table 3 contains the same for the information sources. 11 of the original codes were adopted, 6 were merged to 3 new codes, one was renamed and

6 new codes were introduced to fit the context. 14 codes (10 indicators and 4 sources) were not used in our context and are not given here.

Table 2: Indicators overview: integrability in a novel system [* easy, ** not easy, * out of scope], U = usage of original codes [A=adopted, M=merged, N=new, R=renamed], Q1, Q2 = number of occurrences in questions 1 and 2**

Code	Description	U	Q1	Q2
Academic Workload *	Credit points according to study regulations	A	1	
Assessment Types ** (original: Assessment Tool Types, Course Assessment Tools)	Way the course is examined, e.g. written exam, group project	M	3	3
Course Content *	General description of the content	N	3	
Course Preview **	Preview of what will be covered in the upcoming semester	N		1
Course Requisites *	Recommendations for prior knowledge according study regulations, e.g. other courses or general requisites descriptions	A		1
Course Schedules **	Timetable of offered course	A	5	
Course Type *	Mandatory or elective course	N	1	
Fellow Students' Decisions * (original: Students' Comments)**	Decisions made by fellow students about course enrollments	R	1	
Past Course Grades	Results in previous semesters	A		2
Pending Academic Credits *	Credit points of courses not yet passed	A	1	
Percentage of Passed *	Part of students that passed this course in previous semesters	N		2
Perception of Workload **	Workload perceived by students	A	1	2
Prior Course Syllabus **	Course syllabi in previous semesters	A		1
Program Progress *	Progress according to study regulations	A	1	
Teaching Methodologies **	Teaching methodology used in a course, e.g flipped classroom	A	1	1
Teaching Staff Information **	Information about lecturers, e.g. teaching experience	A	4	1

Table 3: Source overview: U = usage of original codes [A=adopted, M=merged, N=new, R=renamed], Q1, Q2 = number of occurrences in questions 1 and 2

Code	Description	U	Q1	Q2
First Lectures	First 2-3 lectures of a course	N	1	1
Friends and Acquaintances (original: Friends, Other people)		M	5	
Official Evaluation Report (original: Program WhatsApp Group, Student Facebook Group)	Evaluation report provided by the university and filled by students at the end of a course	N		2
Students' Social Media Group	Grouping of various informal social media	M	2	
University Course Catalogue	Study offer and regulations	A	5	

The most important indicators currently used are “course schedules” and “teaching staff information”, while the most important information sources are “friends and acquaintances” and “university course catalogue”. The most often requested indicator is “assessment types” and the most often requested source is "official evaluation report". Based on these results, indicators were identified that can be easily included in the enrollment system (marked with * in Table 2), which will be elaborated in the discussion section.

One notices that four indicators and one source have been mentioned in some groups as an already existing usage (question 1), and as a request for additional information (question 2) in other groups. From the discussions, this can be explained as follows: At the time the enrollment starts, i.e. before the beginning of the semester, only the general information from the university course catalogue is available to students. During especially the first lecture, lecturers provide additional information on assessment, teaching methodologies, and so on. As students have almost 3 weeks after the beginning of the class to cancel their enrollments, for some groups, this is enough. Other groups would like to have this information beforehand, for example in the form of a course preview. Similarly, some groups ask friends about their perceived workload and teaching staff and consider having this information. Other groups would like official information on those items for example, by making parts of the Official Evaluation Report public.

4.2 Part 2

The presentations by the groups and discussions with the students lasted 20 minutes for topic A, and 30 minutes for topics B and C. Needs and concerns as well as the further ideas have been extracted from the students' notes and are listed in Table 4 and Table 5 together with the task in which they were mentioned. "Individual interpretation" refers, for example, to "A1 Performance Comparison" and "B3 Grade predictions". The number of groups in which the item occurred is not shown because only two or three groups worked on the respective tasks in part 2, in contrast to part 1, which was completed by all groups. The improvement ideas will be presented at the end of the discussion section.

Table 4: Needs and Concerns and Related Task (from A1 to C4)

Needs and Concerns	A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4
Equality					x							
Explainability							x	x		x		x
Impact on motivation						x	x					
Individuality					x			x		x		
Individual interpretation	x						x					
Lecturer and course type dependancy			x	x		x	x					

Table 5:Students' Ideas and Related Task (from A1 to C4)

Ideas	A1	A2	A3	A4	B1	B2	B3	B4	C1	C2	C3	C4
Additional recommendations types					x			x	x			
Basic/Expert mode					x							
Dynamic plan for study sequence					x				x		x	
Evaluate individual strengths									x	x		
Other					x		x			x	x	

Needs and concerns are presented in turn, with any conflicting ones listed together:

- Explainability, Individual Interpretation, and Impact on Motivation: Students have little trust in the predictions because not all influencing parameters are known to the system. In their opinion, explanations of how the predictions are made are necessary. Depending on one's performance, grade distributions and predictions may affect students differently. One assumption students make is that performance comparisons are motivating for students with good performance and demotivating for those with poor performance. Another is that students with good performance may become careless. Both assumptions were also mentioned by Bodily and Verbert (2017) and Jivet et al. (2018).
- Equality and Individuality: Students general state that the same information should be available to everyone so that no one feels disadvantaged. However, they have ideas about what additional information could help students with lower grades and also assume that the system cannot take all individual factors into account.
- Lecturer and course type dependency: Students see the lecturers as the biggest factor influencing the grades. Thus, in their opinion, summary statistics of grades in the past and grade prediction should be lecturer dependent, although the course type must also be taken into account: Students consider data exploration and grade predictions to be more helpful for elective courses than for mandatory courses, since for mandatory courses, a choice is possible only if at least two lecturers teach the same course in parallel.

Students' ideas about additional functionalities are mapped against the tasks in which the ideas were discussed in Table 5. A "Dynamic plan for study sequence" – a visualization that brings together courses already taken and a prediction of grades or the best time to take courses that are still open, depending on a maximum number of courses to be taken in the future – could be helpful regarding the individual information demand, individual study condition and supportive in case of a high dropout risk. "Additional recommendations types" refer to recommendations that are not already contemplated, e.g. special recommendations to improve weaknesses or elective courses according to personal interests. "Basic/Expert mode" means a display that can be expanded by the user. "Evaluate individual strengths" refers to one's strengths by competencies, e.g., mathematics, and upcoming courses: if math skills are good, refer to courses with a math component and if math skills are week, refer to the fact that there is not as much math left in the rest of the program. "Other ideas" present further aspects, such as advising sessions or highlighting of career opportunities.

5 DISCUSSION

The result from part 1 provides the basis to answer the research question, identifying indicators that should be included in the enrollment support system. Although all indicators and sources seem reasonable to support course enrollment, there are tradeoffs to be made in realization: some of them can be used more quickly since the data already exist or can be easily obtained (marked with * in Table 2) – all of them are part of the University Course Catalogue. Some items are indicators, that are semester-specific and for which there is currently no structured data, so they cannot be easily included (marked with ** in Table 2). And the last item – Fellow Students' Decision – is out of scope for our future work (marked with *** in Table 2).

Results from part 2 show needs and concerns as well as student ideas to support course enrollment. Contradictions are seen as a challenge: 1. to be equal and fair as well as individual, and 2. to be understandable in order to be trustworthy. Equal information for all is required, but how can e.g. a ranking like "You are in the Top10" of the course motivate the students with good performance and the absence of such a statement not demotivate them? Already the non-appearance of such an evaluation means: "You do not belong to those with the best performance". How to ensure that no learner group benefits more than others as requested by Jivet et al. (2018), considering equity and when low-performing students need more support? To build trust in recommendations, the system must not only be accurate but also explain how the predictions are made so that they are understandable and interpretable. Finally, the SSGD revealed ideas from the students, some of which can be included in the upcoming design phase: A "Dynamic plan for study sequence" was an interesting side finding, that can be an approach to support students. "Other ideas" go beyond of course enrollment support and are outside the scope.

6 CONCLUSIONS AND FUTURE WORK

The purpose of the study was twofold by not only evaluating various functionalities of a novel course enrollment support system through a semi-structured group discussion with students, but also illustrating how to engage students in the design of recommender systems to enhance trust in further implementation stages. This paper includes a description of user involvement as demanded by Bodily and Verbert (2017) and de Quincey et al. (2019), and a set of indicators and sources of information based on the research of Hilliger et al. (2020) and relevant to students at the time of course enrollment. The approach of semi-structured group discussion, which was integrated into a course, is considered to be purposeful: A suitably large number of students was involved to gather relevant aspects for a course enrollment support system. The illustrations of the functionalities with real examples contributed to a lively discussion and provoked both needs and concerns. Future studies will have to look at ways to turn predictions into supportive, understandable, performant, and trusted recommendations, e.g. by using model-agnostic methods. In addition, a quantitative survey supported by an improved prototype and including a larger number of students can provide a better understanding of trust in course recommendations.

REFERENCES

Abdi, S., Khosravi, H., Sadiq, S., & Gasevic, D. (2020). Complementing educational recommender systems with open learner models. Proceedings of the 10th International Conference on Learning Analytics & Knowledge (LAK 2020), 360–365. https://doi.org/10.1145/3375462.3375520

Bodily, R., & Verbert, K. (2017). Trends and issues in student-facing learning analytics reporting systems research. Proceedings of the 7th International Learning Analytics & Knowledge Conference (LAK 2017), 309–318. https://doi.org/10.1145/3027385.3027403

de Quincey, E., Briggs, C., Kyriacou, T., & Waller, R. (2019). Student Centred Design of a Learning Analytics System. Proceedings of the 9th International Conference on Learning Analytics & Knowledge (LAK 2019), 353–362. https://doi.org/10.1145/3303772.3303793

Goncalves, A. L., Carlos, L. M., da Silva, J. B., & Alves, G. R. (2018). Personalized Student Assessment based on Learning Analytics and Recommender Systems. Proceedings of the 3rd International

Conference of the Portuguese Society for Engineering Education (CISPEE 2018), 1–7. https://doi.org/10.1109/CISPEE.2018.8593493

Hilliger, I., De Laet, T., Henríquez, V., Guerra, J., Ortiz-Rojas, M., Zuñiga, M. Á., Baier, J., & Pérez-Sanagustín, M. (2020). For Learners, with Learners: Identifying Indicators for an Academic Advising Dashboard for Students. In C. Alario-Hoyos, M. J. Rodríguez-Triana, M. Scheffel, I. Arnedillo-Sánchez, & S. M. Dennerlein (Eds.), Addressing Global Challenges and Quality Education (pp. 117–130).

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 & Knowledge (LAK 2018), 31–40. https://doi.org/10.1145/3170358.3170421

Ladner, R. E. (2015). Design for user empowerment. Interactions, 22(2), 24–29. https://doi.org/10.1145/2723869

Sarmiento, J., Campos, F., & Wise, A. (2020). Engaging Students as Co-Designers of Learning Analytics. Companion Proceedings of the 10th International Learning Analytics & Knowledge Conference (LAK 2020), 29–32. https://www.solaresearch.org/core/lak20-companion-proceedings/

Wagner, K., Merceron, A., & Sauer, P. (2020a). Erste Untersuchungen zur Notenprognose für ein Kursempfehlungssystem. Proceedings of DELFI Workshops 2020. https://doi.org/10.18420/DELFI2020-WS-112

Wagner, K., Merceron, A., & Sauer, P. (2020b). Accuracy of a Cross-Program Model for Dropout Prediction in Higher Education. Companion Proceedings of the 10th International Learning Analytics & Knowledge Conference (LAK 2020), 744–749. https://www.solaresearch.org/core/lak20-companion-proceedings/

Wilson, A., Watson, C., Thompson, T. L., Drew, V., & Doyle, S. (2017). Learning analytics: Challenges and limitations. Teaching in Higher Education, 22(8), 991–1007. https://doi.org/10.1080/13562517.2017.1332026

Analyzing the Completion Rates of Curricula Using an Iterative Probabilistic Model

Ahmad Slim[1], Georges El-Howayek[2], Elizabeth Bradford[3], Gregory L. Heileman[4], Chaouki T. Abdallah[5], Ameer Slim[6]
Depart. of Computer Science and Mathematics, Lebanese American University, Lebanon[1]
Depart. of Electrical and Computer Engineering, Valparaiso University, Valparaiso, IN[2]
College of Education and Human Science, University of New Mexico, Albuquerque, NM[3]
Depart. of Electrical and Computer Engineering, The University of Arizona, Tucson, AZ[4]
School of Electrical and Computer Engineering, Georgia Tech, Atlanta, GA[5]
Department of Civil Engineering, University of New Mexico, Albuquerque, NM[6]
ahmad.slim@lau.edu.lb, georges.el-howayek@valpo.edu, m.elizabeth.bradford@g-mail.com, heileman@arizona.edu, ctabdallah@gatech.edu, ahs1993@unm.edu

ABSTRACT: Studies show that graduation rates in universities are influenced by the complexity of their curricula. Recent work focused on simulation methods to reveal this kind of correlation. In this paper, we present for the first time a closed-form solution that correlates graduation rates to curricula complexity using a probabilistic approach. We apply our proposed model on a number of curricula with different complexities. The results match those of the simulation methods with an advantage that our closed-form solution is more time-efficient.

Keywords: Curricular analytics, education, student success, probabilistic models, simulation models, dynamic programming.

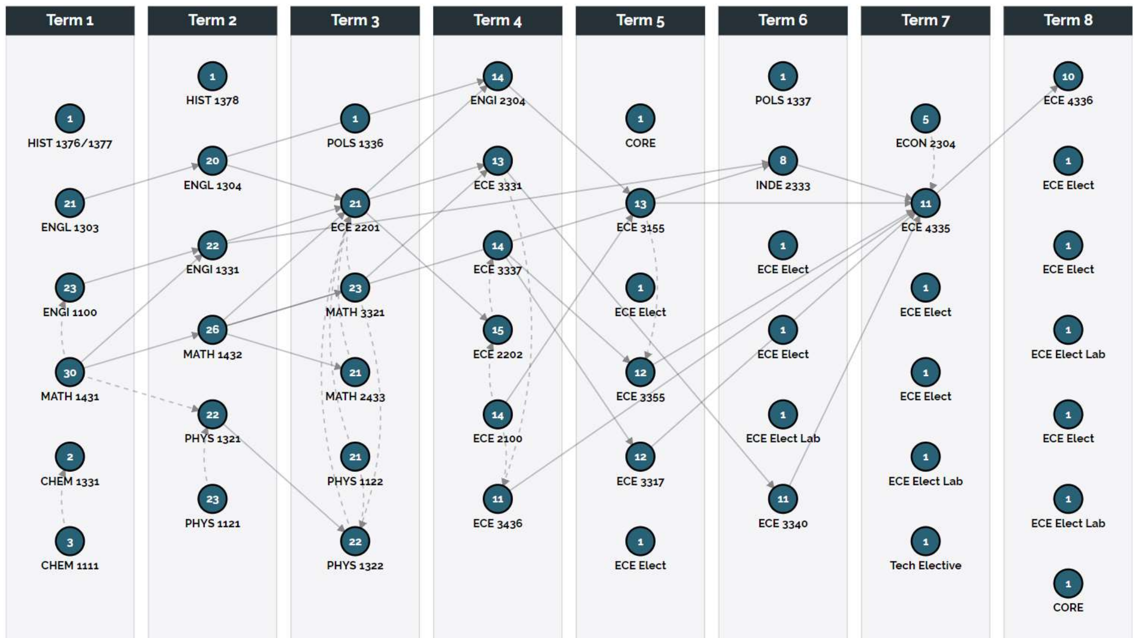
1 INTRODUCTION

Many states are putting special attention for graduation rates at colleges [1]. This stems out from multiple factors that include, but are not limited to, the strong inclination to enhance the quality of college for rating objectives and the growing number of states tying college funding to student performance. Another compelling factor is that a bachelor degree is increasingly becoming essential to thrive in the labor market - making an ethical basis for schools and colleges to accelerate the completion rate pace for the enrolled students. Considering these factors, colleges are increasingly implementing even more sophisticated data driven models on student data to narrow down features influencing persistence and attrition [21, 22]. These contributing features may be broadly categorized as pre-institutional and institutional. The former incorporates features like the pre-college arrangement and financial status, whereas the latter incorporates features that occur when a student is already attending the college (e.g. the approaches, practices, efforts and social interactions, that directly or indirectly influence student success [6, 19]). They concluded that the most essential factor that influences student success is student engagement, noticing that it lies at the intersection between student practices and the aforementioned institutional factors. Many colleges have undertaken these exercises in an endeavor to enhance student success, and have also attempted to expand and bolster student engagement [5, 20]. For example, numerous schools started to track the scholarly advancements of their students more thoroughly and vigorously. Additionally, they started observing the degree to which students engaged in educationally relevant activities, the level of fulfillment with their college experience, and the added value (in terms of learning and aptitudes obtained) of the whole undergraduate experience [7]. A number of institutions reported critical improvement in student accomplishment because of their increased efforts in student engagement, however, other institutions reported minimal improvement. While each of these factors contributes significantly, the most influential parameter for student achievement is degree attainment, and it is very common to discover records of students that procure a degree despite all the odds- they prevail disregarding the chances. It can be unarguably said that for any student, the basic actuality is: if the student passes all the requirements of the degree program, the student would acquire the degree. Consequently, it would be intuitive to think that most of the student success driven interventions mentioned above have to do with supporting the navigation of students through individual requirements tied to a degree program. In fact, the student’s efficiency in navigating through degree program requirements is what matters the most: curricula with less prerequisite dependencies and less courses lead the way to a smoother navigation through the program. In this paper, we study student progress at the most fundamental level, by exploring for the first time- to the best of the authors’ knowledge- the basic properties of a curriculum structure and its relation to graduation rate from a probabilistic approach, quantifying how much curricula complexity affect student achievement.

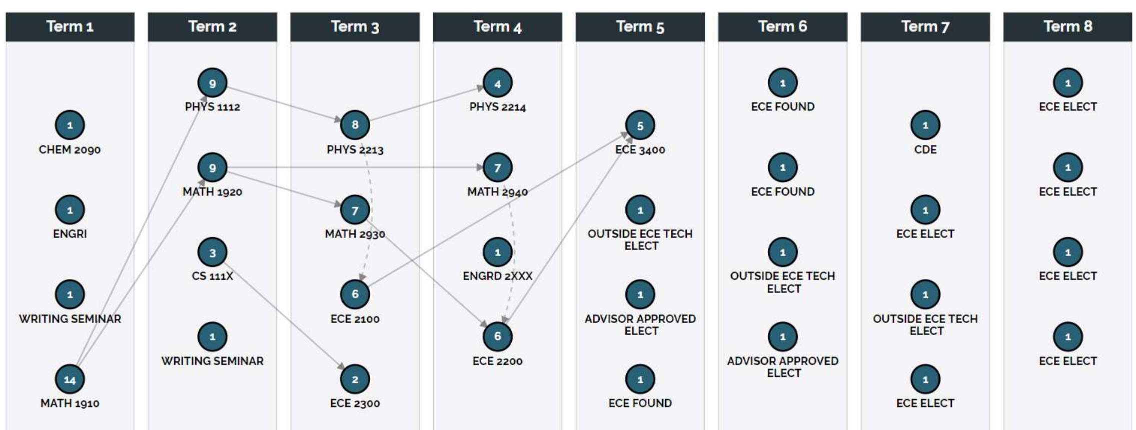
An overview of the paper is as follows. In Section 2, we develop a background for the study of the curriculum by decomposing it along lines related to student success outcomes. In Section 3, we summarize recent work related to curricular analytics done using simulation methods with a particular focus on a Monte Carlo simulation method. Next, in Section 4, we present our proposed model approaching curricular analytics from a probabilistic perspective and comparing it to the simulation method. Finally, Section 5 presents some concluding remarks.

2 BACKGROUND

The analysis of curricula is greatly facilitated by the fact that most academic programs publish their curricula on public websites. This allows those in one academic program to compare their curriculum to those offered by other similar programs. For instance, in Figure 1 we show the undergraduate curriculum of the electrical engineering program provided by two large public institutions in the United States that are similarly accredited by ABET (ABET, 2017). The term-by-term organization of these curricula constitutes the four-year (eight-term) degree plans that students are expected to follow. We have drawn these curricula as graphs, where the vertices represent courses, and the directed edges represent prerequisite arrangements between courses. If there is a directed edge between two courses in the same term, the source vertex is a co-requisite course that may be taken either prior to or at the same time as the course associated with the destination vertex. It is interesting to note that the two programs shown in Figure 1 have identical ABET accreditation. This means that each program satisfies the same eleven ABET program learning outcomes. Thus, from the perspective of ABET, each program is of sufficient quality that the engineers they produce should be prepared to have successful careers. Despite the fact that they have exactly the same accreditation with exactly alike program learning outcomes, it is evident from Figure 1 that these programs have differing structures. In particular, the curriculum in Figure 1a appears far more complex than that in Figure 1b. Students attempting the former must satisfy a much larger number of pre- and co-requisite constraints than those attempting the latter. A logical question that arises instantly is if this would delay students from graduating on time? And if this is true, how can we quantify this delay and how can we determine the correlation between the complexity of a curriculum and graduation rates? By how much do success rates improve with small improvements in curriculum complexity? In the next section, we provide an overview of some of the recent work that investigates these questions in more details and then we present our proposed model.



(a)



(b)

Figure 1: The electrical engineering curricula for the undergraduate program at two large public institutions. Both curricula hold identical ABET accreditation.

3 LITERATURE REVIEW

There has been much research done on curricular analytics investigating the impact of curricula complexity on student performance [16, 15, 18, 13, 11, 12, 2, 14, 3]. Most of these studies use machine learning and statistical models in their analysis. Although less so, there had been a notable effort approaching curricular analytics using simulation models. An example of such studies is the work done by Plotnicki and Garfinke [8]. In their study, they used simulations to design a framework that schedules the courses within a degree program in such a way that allows the majority of students to smoothly traverse through the requirements. A similar work by Schellekens et al. studied the impact of adding more flexibility to students in their programs using simulations [10]. Perhaps the most recent pioneer work in student simulation is presented by Hickman [4]. In his work - named Curricular Analysis and Simulation Library (CASL) simulations - Hickman uses a MonteCarlo method to gauge curricular efficiency by registering a sizable number of virtual students in a degree plan. Particularly, he uses statistical analysis tools to ascertain the passing of each course by each student and then noting down the degree plan completion time for each student. While all of the aforementioned simulations share some similarities with CASL simulations, the work that is mostly related is the study done at San Francisco State University (SFSU) [9]. They implemented a simulation model to determine the effect made on student progress upon changes made to a curriculum. In this work, Saltzman and Roeder used a discrete event simulation to model the flow of students through SFSU's College of Business. In their implementation, they used the school's historic data to simulate students registering for courses. Students would enroll in their chosen courses giving consideration to the prerequisite requirements and the courses' availabilities. Then they either pass or fail based on the historical pass/fail rates of the courses. Using this approach, they were able to determine the effect of restructuring a curriculum on student performance. While this method is closely similar to the CASL approach, they still have some differences. For example, the student behavior in CASL has different assumptions. For instance, the student demand on courses is not based on historical data. CASL also does not allow new incoming classes each term. However, it tracks more student data, supports grade assignments and can be extended to make pass/fail rates determined based on many factors other than historical pass rates data. In this paper, we focus our work on the CASL simulations model and then present its counter version of our proposed closed-form solution model. In the following section, we briefly summarize the procedure followed in the CASL simulations by implementing an experimental trail.

4 CASL'S SIMULATION

To elaborate more about this simulation model, we present one particular experiment and later compare its results to our proposed closed-form solution model. As per this experiment, a set of curricula with a fixed number of courses are considered being offered over a definite number of terms (e.g., four courses offered over two terms). Each curriculum in this experiment is composed of the same courses. Each of these courses has a fixed pass/fail rate. Hence, the structure of the curriculum is the only varying factor in the curricula set. Figure 2 illustrates the structure of each of these curricula. This experiment involves a series of trials. In each trail, a virtual student is simulated flowing through each of these curricula. Then, by the end of all the trails, the completion rate of each course in these curricula is computed in subsequent terms. That is, after running all the trails, the total number of students who passed a certain course in a given curriculum at a given term is computed. It is important to note here that the curriculum completion rates are affected by a number of variables in the simulations and here are some of them: the course pass/fail rate, the number of terms allowed to complete the curriculum, the maximum number of allowed courses in a term, and the characteristics of students who fail to complete the courses. In this particular experiment, all the aforementioned variables are kept constant in each trial. The only varying factor within a given trial was the curriculum structure. This allows to explicitly find the correlation between the structural complexity of the curriculum and graduation rates. The students are not allowed to stop out, i.e., the student would continue to re-attempt the course in subsequent semesters unless successful. It is important to note that this is a highly optimistic assumption that involves determined students who continue to enroll in a course for subsequent terms unless successful.

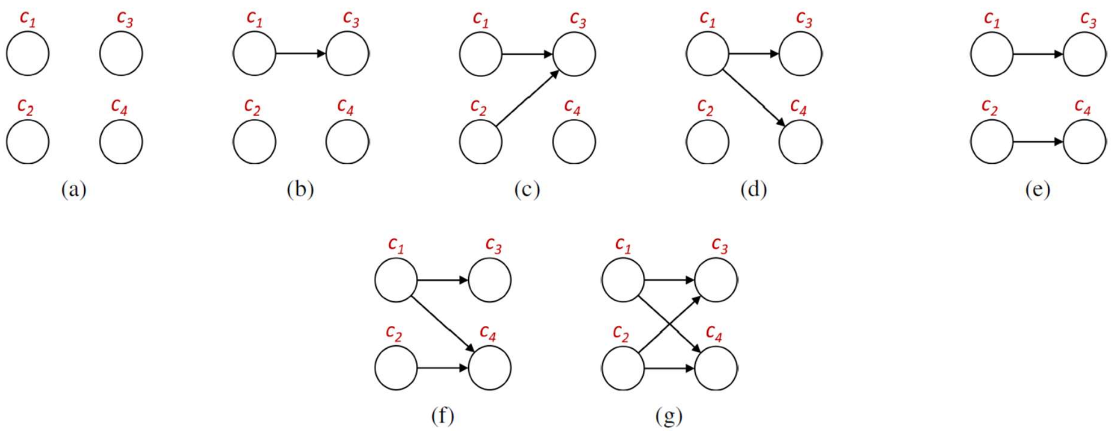


Figure 2: The set of all possible structures for a four-course curriculum. In order, from curricula (a) to (g), the complexities of the curricula using are: 4, 7, 9, 9, 10, 11 and 12.

This leads to higher graduation rates than those observed in real scenarios where students have the option to stop out. However, as noted earlier, the objective of these trials is to relate the structural complexity of the curriculum to the likelihood of completion. The validity of this relationship hold as long as the stop-out behavior is uniform across all the curricula in this experiment. Table 1 shows the graduation rates for each of the curriculum depicted in Figure 2 while keeping the course pass rate for each course at 50%. Particularly, it shows the graduation rates for students at 100% (i.e., term 2), 150% (term 3) and 200% (term 4) time. The results show that the 100%, 150% and 200% graduation rates decrease monotonically as the complexity of the curriculum increases, i.e., going from (a) to (g). This result is indeed intuitive in the sense that the graduation rates are inversely proportional to the structural complexity of the curriculum: as the complexity increases the graduation rates decrease and vice-versa (For more details on how to compute the complexity of a curriculum see [17]). In the next section, we present in more details the probabilistic approach of our proposed closed-form solution version of the CASL simulations and compare the results. The main advantage of our model over simulation models, particularly the CASL simulations, is the efficiency in time. The closed-form solution is more time-efficient than simulations, especially when applied to actual curricula with a large number of courses and prerequisite relationships. Additionally, this tool can be used to instantaneously evaluate curricula and analyze how modifying certain requirements and class prerequisites affects graduation rates.

5 PROBABILISTIC APPROACH

In this section, we derive the probabilities of passing each class in a curriculum as a function of its prerequisite and its passing rate. We first consider a simple scenario where we have a class with a single prerequisite and in later subsections, we gradually increase its complexity and generalize for the case where we have multiple prerequisites.

Table 1: The simulated completion rates for the curricula provided in Figure 2.

(a)	1	2	3	4
c_1	50.0%	75.0%	87.5%	93.75%
c_2	50.0%	75.0%	87.5%	93.75%
c_3	50.0%	75.0%	87.5%	93.75%
c_4	0%	50.0%	75.0%	87.5%
Grad. rate	**0**	**28.13%**	**50.24%**	**72.30%**

(b)	1	2	3	4
c_1	50.0%	75.0%	87.5%	93.75%
c_2	50.0%	75.0%	87.5%	93.75%
c_3	0%	25.0%	50.0%	68.75%
c_4	50.0%	75.0%	87.75%	93.75%
Grad. rate	**0**	**10.55%**	**33.78%**	**56.65%**

(c)	1	2	3	4
c_1	50.0%	75.0%	87.5%	93.75%
c_2	50.0%	75.0%	87.5%	93.75%
c_3	0%	12.5%	34.38%	55.47%
c_4	50.0%	75.0%	87.75%	93.75%
Grad. rate	**0**	**5.75%**	**23.23%**	**45.71%**

(d)	1	2	3	4
c_1	50.0%	75.0%	87.5%	93.75%
c_2	50.0%	75.0%	87.5%	93.75%
c_3	0%	25.0%	50.0%	68.75%
c_4	0%	25.0%	50.0%	68.75%
Grad. rate	**0**	**3.52%**	**19.25%**	**41.54%**

(e)	1	2	3	4
c_1	50.0%	75.0%	87.5%	93.75%
c_2	50.0%	75.0%	87.5%	93.75%
c_3	0%	25.0%	50.0%	68.75%
c_4	0%	25.0%	50.0%	68.75%
Grad. rate	**0**	**3.52%**	**19.25%**	**41.54%**

(f)	1	2	3	4
c_1	50.0%	75.0%	87.5%	93.75%
c_2	50.0%	75.0%	87.5%	93.75%
c_3	0%	25.0%	50.0%	68.75%
c_4	0%	12.5%	34.38%	55.47%
Grad. rate	**0**	**1.76%**	**13.24%**	**33.52%**

(g)	1	2	3	4
c_1	50.0%	75.0%	87.5%	93.75%
c_2	50.0%	75.0%	87.5%	93.75%
c_3	0%	12.5%	34.38%	55.47%
c_4	0%	12.5%	34.38%	55.47%
Grad. rate	**0**	**1.58%**	**9.10%**	**27.04%**

5.1 Passing Probability with no Prerequisite

Let c_i be a class in a particular curriculum and assume that the course pass rate is α_i. Let X_i be a discrete random variable that represents the semester index of class c_i that is passed. The trivial case to consider is when class c_i has no prerequisite. In that case, the probability of passing class c_i in the first semester is $P(X_i = 1) = \alpha_i$. Similarly, the probability of passing class c_i in the second semester is $P(X_i = 2) = \alpha_i(1 - \alpha_i)$ where $(1 - \alpha_i)$ represents the probability of failing class c_i in the first semester.

Also, the probability to pass class c_i in or before the second semester is $P(X_i \leq 2) = 2\alpha_i - \alpha_i^2$. In general, the pass rate of class i at semester $n > 1$ is

$$P(X_i = n) = \alpha_i(1 - P(X_i \leq n - 1)) \tag{1}$$

and the cumulative probability is

$$P(X_i \leq n) = P(X_i \leq n - 1)(1 - \alpha_i) + \alpha_i \tag{2}$$

5.2 Passing Probability with one Prerequisite

The second simple scenario to consider is when class c_i has only one prerequisite class, c_a. Obviously, in the case when the cumulative passing probability of class c_a at semester $n - 1$ is zero (i.e. when $P(X_a \leq n - 1) = 0$), the passing probability of class c_i at a semester n is also zero (i.e. $P(X_i = n) = P(X_i \leq n) = 0$). In other words, the class c_i cannot be passed without passing its prerequisite c_a. However, when the cumulative passing probability of the prerequisite class c_a is non-zero, the probability of pass c_i in semester n is

$$P(X_i = n) = \alpha_i[P(X_a \leq n - 1) - P(X_i \leq n - 1)] \tag{3}$$

The terms between the brackets on the right-hand side of the latter equation represent the cumulative probability to pass the prerequisite class c_a in any semester before n without passing c_i. Multiplying the result by α_i, we can determine the probability of passing c_i in semester n. Consequently, the recursive equation of the cumulative probability is

$$P(X_i \leq n) = P(X_i \leq n - 1) + P(X_i = n) \tag{4}$$

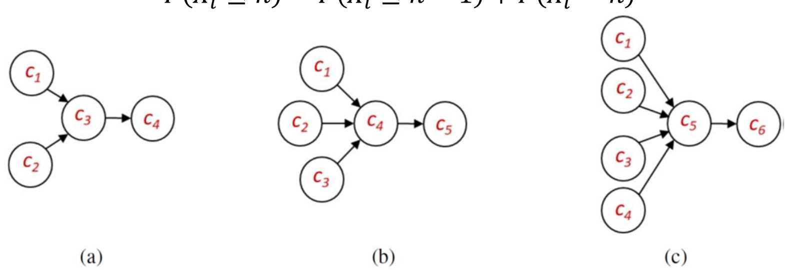


Figure 3: Different scenarios for classes with (a) two pre-requisites, (b) three pre-requisites, and (c) four pre-requisites.

5.3 Passing Probability with two Prerequisites

When considering more than one class as prerequisites, it is required to determine the probability of passing all the prerequisite classes (i.e., the intersection probability). In this section, we show how to compute the intersection probability of two prerequisite classes a and b. Then, using (3) and (4) and replacing $P(X_a \leq n - 1)$ by the intersection probability, we can determine $P(X_i = n)$ and $P(X_i \leq n)$. The intersection probability $P(X_a \leq n \cap X_b \leq n)$ can be found using the following recursive equation:

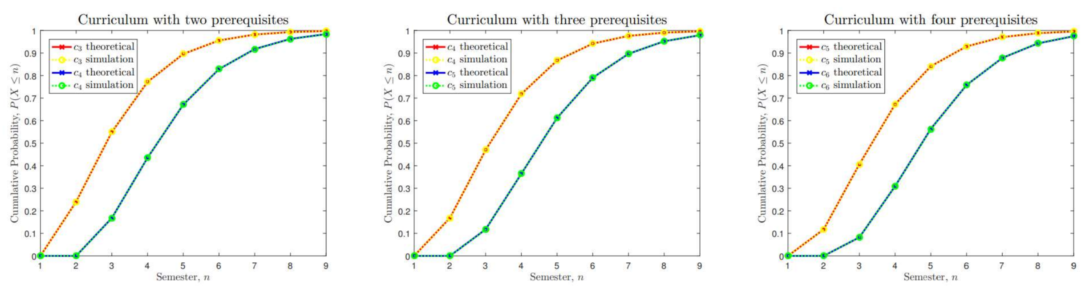
$$\begin{aligned} P(X_a \leq n \cap X_b \leq n) \\ &= P(X_a \leq n - 1 \cap X_b \leq n - 1) + P(X_a \leq n - 1 \cap X_b = n) \\ &+ P(X_a = n \cap X_b \leq n - 1) + P(X_a = n \cap X_b = n) \end{aligned} \tag{5}$$

The four terms in the right-hand side represent the probabilities of the four disjoint events that cover all the possible cases that a student can pass classes a and b in semester n or before. The first term represents the probability of the event of passing classes a and b in semester $n - 1$ or before. The second term is the probability of passing class b in semester n while class a is passed before that semester (in semester $n - 1$ or before). Conversely, the third term is the probability of passing class a

in semester n and class b before that. Finally, the last term is the probability of passing both classes a and b in semester n. Each of these terms can be found as follows:

- The 1^{st} term represents the recursive term for $n \leq 1$. The initial condition is zero.
- The 2^{nd} term is found by first considering the probability of passing class a in semester $n-1$ or before ($P(X_a \leq n-1)$) excluding the probability of passing both classes in semester $n-1$ or before ($P(X_a \leq n-1 \cap X_b \leq n-1)$). This equivalent to the probability of passing class a and not yet passing class b. Finally, by multiplying it by the passing rate α_b, we determine the probability of passing class a in any semester before n and class b in semester n. i.e.,
$$P(X_a \leq n-1 \cap X_b = n) = \alpha_b \times (P(X_a \leq n-1) - P(X_a \leq n-1 \cap X_b \leq n-1)) \tag{6}$$
- The 3^{rd} term represents the probability of passing class a in semester n and class b in or before semester $n-1$. Similarly to the second term, it can be found as follows:
$$P(X_a = n \cap X_b \leq n-1) = \alpha_a \times [P(X_b \leq n-1) - P(X_a \leq n-1 \cap X_b \leq n-1)] \tag{7}$$
- The 4^{th} term is the probability of passing both classes at term n. It is equal to the probability of passing the prerequisite class of a and b (denoted by p) before semester n minus the probability of passing either class a or class b (the union probability) multiplied by the passing rate of class a and b. Therefore,
$$P(X_a = n \cap X_b = n) = \alpha_a \times \alpha_b \times [P(X_p \leq n-1) - P(X_a \leq n-1 \cup X_b \leq n-1)] \tag{8}$$
where the union probability can be found using the union rule:
$$\begin{aligned} P(X_a \leq n-1 \cup X_b \leq n-1) \\ = P(X_a \leq n-1) + P(X_b \leq n-1) - P(X_a \leq n-1 \cap X_b \leq n-1)) \end{aligned} \tag{9}$$

Similarly, those results can be extended for more than two prerequisite classes.



(a) The cumulative probability of passing classes c_3 and c_4 as a function of semester n for the scenario depicted in Figure 3a

(b) The cumulative probability of passing classes c_4 and c_5 as a function of semester n for the scenario depicted in Figure 3b

(c) The cumulative probability of passing classes c_5 and c_6 as a function of semester n for the scenario depicted in Figure 3c

Figure 4

6 EXPERIMENTAL RESULTS

To validate our proposed probabilistic approach, we calculated the probability of three different scenarios and compared the results to those of the CASL simulations. In this simulation, the passing probability for each class, α_i, is set to 0.7 and the cumulative probability is calculated for the first 9 semesters. The first scenario considered is depicted in Figure 3a. In this scenario, we considered the case of classes with two prerequisites. Using (5), the intersection probability of class c_1 and c_2 was calculated. The result was later used to calculate the cumulative probability of classes c_3 and c_4 using (3) and (4). The results of the calculation are shown in Figure 4a. As expected, the cumulative probability for c_3 starts in the second semester with a probability of 24% and gradually increases to converge at 100%. Similarly, the cumulative probability of c_4 starts in the third semester with a probability lower than c_3 (17.1%). In both cases, the probabilistic approach calculations match perfectly with the CASL simulations. For the second scenario depicted in Figure 3b, we expand the closed-form expressions of the intersection probability shown in (5) to consider three classes scenario. The results were later used to calculate the cumulative probability of c_4 and c_5 using (3) and (4). The calculated and simulated probabilities are shown in Figure 4b. As the number of the prerequisites increases, the cumulative probability decreases. The cumulative probability for c_4 starts in the second semester with a value of 16.87% and gradually converges to 100%. Similarly, the cumulative probability of c_5 starts in the third semester with a value of 11.8% and gradually converges to 100%. For the last scenario, we consider the case with four prerequisites (Figure 3c). After determining the intersection probability of the four prerequisite classes, the cumulative probabilities of c_5 and c_6 are calculated. The cumulative probabilities of c_5 and c_6 start at 11.8% and 8.2%, respectively.

7 CONCLUSION

The graduation rates in universities are influenced by the complexity of their curricula. In this paper, we present a closed-form equations to compute the completion rates of curricula. Recent work focused mainly on simulation methods to compute the completion rates. However, these methods proved not to be scalable when applied to actual curricula with a large number of courses and prerequisites. Our work fills in this gap by presenting a more time-efficient model using a probabilistic approach. Moreover, the model presented here can be used as a tool to study the effects of changes in a curriculum on the graduation rate. We applied our proposed model on different curricula. The results of our model match exactly those of the simulation models. In the future, we intend to extend our work to include more realistic scenarios that simulates actual institutional variables such as the maximum number of courses allowed per semester, the number of times a student is allowed to take a course, stop-outs, etc.

REFERENCES

[1] *Performance funding for higher education*. National Conference of State Legislatures, Feb. 2012.

[2] G. L. Heileman, C. T. Abdallah, A. Slim, and M. Hickman. Curricular analytics: *A framework for quantifying the impact of curricular reforms and pedagogical innovations*. CoRR, abs/1811.09676, 2018.

[3] G. L. Heileman, M. Hickman, A. Slim, and C. T. Abdallah. *Characterizing the complexity of curricular patterns in engineering programs*. In 2017 ASEE Annual Conference and Exposition, Columbus, Ohio, 2017. ASEE Conferences.

[4] M. S. Hickman. *Development of a curriculum analysis and simulation library with applications in curricular analytics*. Master's thesis, The University of New Mexico, 2017.

[5] G. D. Kuh, J. Kinzie, J. A. Buckley, B. K. Bridges, and J. C. Hayek. *What matters to student success: A review of the literature*. Technical report, National Postsecondary Education Cooperative, U.S. Department of Education, Commissioned Report for the National Symposium on Postsecondary Student Success: Spearheading a Dialog on Student Success, 2006.

[6] G. D. Kuh, J. Kinzie, J. H. Schuh, and E. J. Whitt. *Student Success in College: Creating Conditions That Matter*. Jossey-Bass, San Francisco, CA, 2010.

[7] C. Moore and N. Shulock. *Student progress toward degree completion: Lessons from the research literature*. Technical report, The Institute for Higher Education Leadership & Policy, California State University, Sacramento, 2009.

[8] W. Plotnick and R. Garfinkel. *Scheduling academic courses to maximize student flow: A simulation approach*. 20(4):193–199, 1986.

[9] R. M. Saltzman and T. M. Roeder. Simulating student flow through a college of business for policy and structural change analysis. *Journal of the Operational Research Society*, 63(4):511–523, 2012.

[10] A. Schellekens, F. Paas, A. Verbraeck, and J. J. G. van Merriënboer. Designing a flexible approach for higher professional education by means of simulation modelling. *Journal of the Operational Research Society*, 61(2):202–210, Feb 2010.

[11] A. Slim, G. L. Heileman, W. Al-Doroubi, and C. T. Abdallah. *The impact of course enrollment sequences on student success*. In 2016 IEEE 30th International Conference on Advanced Information Networking and Applications (AINA), pages 59–65, March 2016.

[12] A. Slim, G. L. Heileman, M. Hickman, and C. T. Abdallah. *A geometric distributed probabilistic model to predict graduation rates*. In 2017 IEEE Cloud Big Data Computing (CBDCom), pages 1–8, Aug 2017.

[13] A. Slim, G. L. Heileman, J. Kozlick, and C. T. Abdallah. *Predicting student success based on prior performance*. In Proceedings of the 5th IEEE Symposium on Computational Intelligence and Data Mining, Orlando, FL, 2014.

[14] A. Slim, D. Hush, T. Ojha, C. T. Abdallah, G. L. Heileman, and G. El-Howayek. *An automated framework to recommend a suitable academic program, course and instructor*. In Proceedings of the Fifth International Conference on Big Data Computing Service and Applications (BigDataService), San Francisco, CA, USA, 2019.

[15] A. Slim, J. Kozlick, G. L. Heileman, and C. T. Abdallah. *The complexity of university curricula according to course cruciality*. In Proceedings of the 8th International Conference on Complex, Intelligent, and Software Intensive Systems, Birmingham City University, Birmingham, UK, 2014.

[16] A. Slim, J. Kozlick, G. L. Heileman, J. Wigdahl, and C. T. Abdallah. *Network analysis of university courses*. In Proceedings of the 6th Annual Workshop on Simplifying Complex Networks for Practitioners, Seoul, Korea, 2014.

[17] A. Slim, J. Kozlick, G. L. Heileman, J. Wigdahl, and C. T. Abdallah. *Network analysis of university courses*. In Proceedings of the 6th Annual Workshop on Simplifying Complex Networks for Practitioners, Seoul, Korea, 2014.

[18] A. H. Slim, G. L. Heileman, J. Kozlick, and C. T. Abdallah. *Employing markov networks on curriculum graphs to predict student performance*. In Proceedings of the 13th International Conference on Machine Learning and Applications, Detroit, MI, 2014.

[19] V. Tinto. *Dropout from higher education: A theoretical synthesis of recent research*. Review of Educational Research, 1(45):89–125, 1975.

[20] V. Tinto. *Leaving College: Rethinking the Causes and Cures of Student Attrition*. University of Chicago Press, Chicago, IL, 1987.

[21] A. Venezia, P. M. Callan, J. E. Finney, M. W. Kirst, and M. D. Usdan. *The governance divide: A report on a four-state study on improving college readiness and success*. Technical report, The National Center for Public Policy and Higher Education, San Jose, CA, Sep. 2005.

[22] L. Zhang. Does state funding affect graduation rates at public four-year colleges and universities? *Educational Policy*, 23(5):714–731, Sep. 2009.

Towards a Philosophical Framework for Learning Analytics (POLA@LAK21)

Pablo Munguia
Flinders University
pablo.munguia@flinders.edu.au

Andrew Gibson
Queensland Institute of Technology (QUT)
andrew.gibson@qut.edu.au

ABSTRACT

This workshop aims at discussing and sharing ideas to help construct a philosophical framework that learning analytics needs as a field. This workshop is the first step towards the development of a philosophical approach to help practitioners collaborate, interrogate, and develop this foundation. The workshop is a half-day event. Participants will be invited to submit a brief position paper for review in advance of the workshop. During the event there will be brief presentations of these papers followed by collaborative activities to create robust, but intellectually stimulating and constructive conversations. The workshop will be synthesized via a multi-authored publication summarizing the points discussed to share with the broader field.

KEYWORDS: Philosophical framework, theories, learning analytics as a discipline.

1. BACKGROUND

Learning analytics is maturing as a discipline. Yet as practitioners from diverse backgrounds bring their varying skill sets, they also bring their own disciplinary based philosophies to the field, which at times can introduce confusion or even dissonance. We argue that there is a need to initiate a conversation about how learning analytics should be philosophically grounded and argue for a philosophical framework for the field. We believe a workshop that facilitates this discussion is timely given the questions posed for the 11th Annual Learning Analytics and Knowledge conference.

We can assume that all learning analytics practitioners share the same ultimate passion and goal, to find qualitative and quantitative ways to improve the learning experience for learners. However, the pathways we take are highly diverse, which can result in resistance when trying to bridge discourse between disciplines. We suggest that learning analytics as a transdisciplinary field would benefit from building on a philosophical construct to allow improved collaboration and uptake by academics and institutions.

The lack of a philosophical framework for learning analytics has significant ramifications which may prevent the field from maturing. For example, it causes confusion when trying to explain what is and is not in scope for the field. Further, the absence of a philosophical framework slows the field down when attempting to test, experiment and scale up ideas and methods.

We are still debating at length the ethics surrounding the discipline (e.g., Corrin et al., 2019, Ferguson 2019, Kitto and Knight 2019, West et al. 2020), which a philosophical framework would help resolve. Selwyn's (2020) provocations express the need to dig deep and assess whether the current direction of learning analytics is indeed what we want for the field. More importantly, Selwyn questions what is actually needed in society and what is missing from our background disciplines when moving into this transdisciplinary space. Finally, such a philosophical framework would allow for the creation of momentum, as the field is reaching a critical turning point: it is needs to move beyond a few practitioners working in isolation or practicing in few classrooms to institutional or national plans to adopt and follow ethical use of learner data for pedagogical purposes. This last point is being made frequently (e.g., Ferguson 2012, Selwyn 2020 West et al. 2020), and while a recent survey showed that institutions are willing (Tsai and Gasevic 2017), when attempting to put in place these methods, we often fail (Ferguson 2012, Buerck 2014, Munguia et al. 2020).

2. ORGANIZATIONAL DETAILS

WORKSHOP TYPE: Interactive Workshop session**.**

WORKSHOP SIZE: we are targeting 12-15 Participants. With 5-10 key submissions.

DURATION: half a day.

TYPE OF PARTICIPATION: mixed participation (interested delegates may submit a short paper, see below).

EXPECTED ACTIVITIES: short presentations by participants that submitted papers, discussion groups, working on a publication authored by interested participants.

3. OBJECTIVES:

We are creating a sharing and collaborative workshop for two groups of people:

1. those that have an a priori contribution to make about a philosophical framework of learning analytics as a field; and
2. those who have a more general interest in the topic and would like to engage with the ideas proposed by others.

The workshop is designed to meet the following objectives:

1. Initiate a conversation around developing a philosophical framework for learning analytics
2. Provide a forum of friendly critique for existing ideas
3. Present the discussion of ideas in a form that can be disseminated to the wider community.

In order to meet objective 3, we intend that an output of the workshop will be a synthetic paper harvesting participants' input. Our intention is to submit this paper for publication in the JLA.

3. POSITION PAPER SUBMISSION

Workshop participants can submit a paper for the workshop. The length should be a maximum of 3 pages. Authors should focus on answering the question:

What philosophical ideas should be considered as a foundation for the field of learning analytics and how might practitioners in the field engage with them?

Deadline for submission of contributions is 9 February 2021.

WORKSHOP FORMAT

Time	Activity
1 hour	Contributors present key ideas lightning talk style (4-5 mins each). Other participants write PMI (Plus/Minus/Interesting) short reactions on post-it notes which are grouped.
1-1.5 hours	Group break-out around key philosophical ideas presented - 'birds of a feather' style. Groups engage in dialogue around (a) the potential role of the idea in LA, (b) the value to ALL LA stakeholders, (c) how the idea might advance or secure the field moving forward.
1.5-2 hours	Bring ideas together from all groups and begin writing synthesis paper. Break out in smaller groups as necessary.

REFERENCES

Buerck, J. P. (2014). A Resource-Constrained Approach to Implementing Analytics in an Institution of Higher Education: An Experience Report. Journal of Learning Analytics, 1(1), 129-139. https://doi.org/10.18608/jla.2014.11.7

Corrin, L., Kennedy, G., French, S., Buckingham Shum S., Kitto, K., Pardo, A., West, D., Mirriahi, N., & Colvin, C. (2019). The Ethics of Learning Analytics in Australian Higher Education. A Discussion Paper.https://melbourne-cshe.unimelb.edu.au/research/research-projects/edutech/the-ethical-use-of-learning-analytics.

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4, 304–317. https://doi.org/10.1504/IJTEL.2012.051816

Ferguson, R. (2019). Ethical Challenges for Learning Analytics. *Journal of Learning Analytics*, *6*(3), 25–30. https://doi.org/10.18608/jla.2019.63.5

Kitto, K. and Knight, S. 2019. Practical ethics for building learning analytics. Br J Educ Technol, 50: 2855-2870. doi:10.1111/bjet.12868

Munguia, P., Brennan, A., Taylor, S., & Lee, D. (2020). A learning analytics journey: Bridging the gap between technology services and academic need. The Internet and Higher Education. 46: 100744.

Selwyn, Neil. "Re-Imagining 'Learning Analytics' … a Case for Starting Again?" *The Internet and Higher Education* 46 (July 2020): 100745. https://doi.org/10.1016/j.iheduc.2020.100745.

Tsai, Y. S., & Gašević, D. (2017). Learning analytics in higher education—challenges and policies: A review of eight learning analytics policies. Proceedings of the 7th International Learning Analytics & Knowledge Conference (pp. 233–242).

West, D., Luzeckyj, A., Toohey, D., Vanderlelie, J., & Searle, B. 2020. Do academics and university administrators really know better? The ethics of positioning student perspectives in learning analytics. Australasian Journal of Educational Technology, 36(2), 60-70. https://doi.org/10.14742/ajet.4653

Understanding LAK by Understanding its Philosophical Paradigms

Ryan S. Baker
University of Pennsylvania
rybaker@upenn.edu

Dragan Gasevic
Monash University
dragan.gasevic@gmail.com

Keywords: Philosophy of learning analytics, Richard McKeon, paradigms for learning analytics

1 EXTENDED ABSTRACT

There is much debate about what properties a paper in LAK or related conferences should have. However, despite reviewers often trying to decide if a paper is "an EDM paper" or "a LAK paper", there has been limited systematic consideration of how these fields differ (but see Siemens & Baker, 2012, and topic analysis in Chen et al., 2020).

We propose that the difference in expectations between communities comes from the predominant (often unspoken) philosophical orientations of their members. The 20th-century philosopher Richard McKeon wrote about four philosophical schools of thought dating back to Plato and Aristotle (McKeon, 1966): the entitative/reductionist (Atomist/Democritus), ontological/dialectical (Platonic), existentialist (Sophist/Protagoras), and essentialist (Aristotelian) schools of thought. Reductionism involves understanding complex phenomena by breaking down those phenomena into their constituent components and then analyzing the relationship between those components. The ontological/dialectical school of thought adopts the goal of understanding phenomena as wholes, where components cannot be properly understood without understanding the whole system. Existentialism views reality as fundamentally individually constructed and therefore asserts that phenomena should be understood as the participants themselves understand them and that these understandings are irreducibly valid. This viewpoint's opposite, Essentialism*,* states that meaning is inherent in the universe. This viewpoint is seen in perspectives that argue for the "unreasonable effectiveness of data" as justification for rejecting interpretable modeling methods (Halevy, Norvig, & Pereira, 2009), where direct modeling of reality is seen as sufficient and no attempt at theory or explanation is needed (or, indeed, desired). The design theorist Dick Buchanan argued (in largely unpublished lectures) that most designers prefer to work in one of these paradigms, and that work from other perspectives often seems confusing or perhaps even intentionally incomprehensible or negative.

Siemens and Baker (2012) mapped learning analytics to two of these paradigms, arguing that most of the work in educational data mining had a "stronger emphasis on reducing to components and analyzing individual components and relationships between them" (*ibid.*, p. 253) (i.e. entitative, reductionist) whereas most of the work in learning analytics had a "stronger emphasis on understanding systems as wholes, in their full complexity" (*ibid.*, p. 253) (i.e. ontological, dialectical). Work from a more essentialist paradigm tended to fare relatively poorly at EDM and LAK in these early years, and the emergence of the ACM Learning @ Scale conference provided a home for more essentialist work. With the success of ACM Learning @ Scale, and the popular movement within machine learning towards algorithms that do not attempt to be scrutable, there was a movement towards more serious consideration of prediction without comprehension within EDM. This movement was matched by a corresponding movement of entitative work into LAK, with many researchers who had previously published at AIED or EDM beginning to publish their work at LAK. More existentialist learning analytics research often appeared outside these conferences until the emergence of the International Conference on Quantitative Ethnography (followed by a large increase in the use of quantitative ethnography methods at LAK in 2020).

When a scientific field contains more than one competing intellectual paradigm, there is often a push for one of the paradimgs to "win" and take over the field entirely, as depicted in Kuhn's (1962) classic book *The Structure of Scientific Revolutions*. We hope to argue that different modes of thought are natural, are positive, and are good for the field. The complex challenges that learning analytics poses to us as researchers and practitioners (cf. Baker, 2019; Pelanek, 2020) are too large to be entirely resolved by any of these four paradigms. There needs to be greater collaboration across researchers from different intellectual paradigms -- a move towards *inter-paradigmatic* work in addition to the inter-disciplinarity that already characterizes our field.

REFERENCES

Baker, R.S. (2019) Challenges for the future of educational data mining: The Baker learning analytics prizes. *Journal of Educational Data Mining*, *11*(1), 1-17.

Chen, G., Rolim, V., Mello, R. F., & Gašević, D. (2020). Let's shine together! a comparative study between learning analytics and educational data mining. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 544-553).

Halevy, A., Norvig, P., & Pereira, F. (2009). The unreasonable effectiveness of data. *IEEE Intelligent Systems*, *24*(2), 8-12.

Kuhn, T.S. (1962). *The structure of scientific revolutions* (1st ed.). University of Chicago Press.

McKeon, R. (1966). Philosophic semantics and philosophic inquiry. In Z. K. Mckeon (Ed.), *Freedom and history and other essays: An introduction to the thought of Richard McKeon* (pp. 242-256). The University of Chicago Press.

Pelánek, R. (2020). Learning analytics challenges: trade-offs, methodology, scalability. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 554-558).

Siemens, G., Baker, R.S.J.d. (2012) Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*.

Building shared understandings of learning analytics worldviews: A role for structured dialogue

Kristine Lund
Laboratoire ICAR, ENS de Lyon, CNRS, Université Lumière Lyon 2
Kristine.Lund@ens-lyon.fr

Stephen Crowley
Boise State University
stephencrowley@boisestate.edu

Michael O'Rourke
Michigan State University
orourk51@msu.edu

Keywords: epistemology, research assumptions, team science, research community, structured dialogue

1 EXTENDED ABSTRACT

Science makes progress through teamwork and when a research field focuses on complex problems with societal impact, team-members need diverse skill sets in order to bring the necessary knowledge and practical experience to the table (Hall, et al., 2012). But a team whose members have diverse knowledge and experience often run into obstacles once they begin working together. They may not make their assumptions explicit about the problem they are trying to solve and they may have differing beliefs and values about particular aspects of the problem (Lund, Rosé, Suthers, & Baker, 2013). They often have different ideas about how to do research, or what constitutes the perimeter of their own activity or those with whom they work. If team-members do not understand these differences before they begin a project together, such differences may emerge at inopportune times, produce conflicts, and temporarily halt progress, or more seriously, even end the collaboration.

These difficulties occur at the level of a research team, but are also relevant for a community of practice (Lund, Jeong, Grauwin, & Jensen, 2020) and *Learning Analytics* is a case in point (Balacheff & Lund, 2013). On-line community discussion[1] has revealed a number of diverse assumptions by community members regarding many topics, some of which are below. All of them could result in roadblocks for research teams pursuing a shared objective (Rosé & Lund, 2013) and/or for research communities founded on different types of shared knowledge on which they depend:

- Differing preconceived notions regarding relations with stakeholders outside of academia;
- Partial alignment of the goals for engineering versus the goals of research;

[1] learninganalytics@googlegroups.com

- Differences in the value attributed to "outliers", something that may be ignored in research, but that can be a matter of life or death in engineering

- Disagreement on the definition, competencies, roles of or even the existence of a "learning engineer" in a field called "Learning Engineering";

- Differing opinions on the fundamental differences of scope between the sister communities Learning Analytics, Educational Data Mining, and Learning Sciences, as well as in relation to broader umbrella community terms such as Learning Informatics[2];

- Disagreement on whether techno-solutionism is embraced by researchers or rather occurs only at the vendor level.

This proposal suggests a methodical way to bring such disagreements to light, confront them, hash them out, and thereby improve communicative and collaborative capacity within heterogeneous research teams (O'Rourke, & Crowley, 2013; Crowley & O'Rourke, 2020). The proposal also specifically addresses recognizing differences and building common ground in a community of research. The method is based on the Toolbox Dialogue Initiative (e.g. Hubbs, O'Rourke, & Orzack, 2020). Toolbox workshops[3] help groups discover and examine perspectives by using questionnaires and structured dialogue that reveal attitudes, views, values, and beliefs. Workshop facilitators ask participants to rank a set of *probing statements* on a Likert scale[4] and then use an app to collect the responses on a server and form discussion groups based on the responses. The probing statements are claims that are designed to help a participant see his/her biases and through subsequent discussion, move toward recognizing unacknowledged differences with other group members. Sample *probing statements*[5] in Learning Analytics for which participants would position themselves on a Likert scale could include the following:

1. "We can't solve a problem irrespective of the context in which it is used."

2. "Anything that is complex cannot be engineered."

3. "Education is broken and it should and can be fixed with technology" (Teräs, Suoranta, &Teräs, 2020).

4. "Engineering is about making things; science is about understanding things"

5. "Learning engineers do not engineer learning, but rather engineer learning systems"

---

[2] http://simon.buckinghamshum.net/2020/09/why-learning-informatics

[3] Center for Interdisiciplinarity http://c4i.msu.edu/

[4] (1) Strongly disagree; (2) Disagree; (3) Neither agree nor disagree; (4) Agree; (5) Strongly agree.

[5] learninganalytics@googlegroups.com

6. "We can't reduce the real problem of learning down to engineerable problems"

7. "Framing something as an engineering problem implies that it can be controlled, predicted, or managed in the same way that we can manage a fuel pump."

Given that we are submitting to the workshop on Philosophy of Learning Analytics and that it is designed to initiate a conversation around developing a philosophical framework for learning analytics, we propose to collaboratively develop a more substantial set of probing statements that could be used in future instances of a full Toolbox workshop. Developing such probing statements and discussing them seems crucial for moving the field forward.

REFERENCES

Crowley, S., O'Rourke, M. (2020). Communication failure and cross-disciplinary research. In G. Hubbs, M. O'Rourke, and S. H. Orzack (Eds.), *The Toolbox Dialogue Initiative: The Power of Cross-Disciplinary Practice* (pp. 1–16). Boca Raton, FL: CRC Press.

Balacheff, N. & Lund, K. (2013). Multidisciplinarity vs. Multivocality, the case of "Learning Analytics". *LAK 2013 - International Conference on Learning Analytics and Knowledge*, Apr 2013, Leuven, Belgium. pp. 5-13, ⟨10.1145/2460296.2460299⟩.

Hall, K. L., Vogel, V.L., Stipelman, B.A., Stokols, D., Morgan, G., & Gehlert, S. (2010). A four-phase model of transdisciplinary team-based research: goals, team processes, and strategies. *TBM* 2012;2:415–430, doi: 10.1007/s13142-012-0167-y.

Hubbs, G., O'Rourke, M., Orzack, S. H. (Eds.). (2020). *The Toolbox Dialogue Initiative: The Power of Cross-Disciplinary Practice.* Boca Raton, FL: CRC Press.

Lund, K., Jeong, H., Grauwin, S. and Jensen, P. (2020). Research in Education Draws Widely From the Social Sciences and Humanities. *Front. Educ*. 5:544194. doi: 10.3389/feduc.2020.544194.

Lund, K., Rosé, C. P., Suthers, D. D., and Baker, M. (2013). Epistemological encounters in multivocal settings, in D. D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, N. Law (Eds.), *Productive Multivocality in the Analysis of Group Interactions*, (pp. 659–682). New York, NY: Springer.

O'Rourke, M., & Crowley, S. J. (2013). Philosophical intervention and cross-disciplinary science: The story of the Toolbox Project. *Synthese*, 190(11), 1937–1954.

Rosé, C. P., and Lund, K. (2013). Methodological pathways for avoiding pitfalls in multivocality, in D. D. Suthers, K. Lund, C. P. Rosé, C. Teplovs, and N. Law (Eds.) *Productive Multivocality in the Analysis of Group Interactions*, (pp. 613–637). New York, NY: Springer.

Teräs, M., Suoranta, J., Teräs, H. et al. (2020). Post-Covid-19 Education and Education Technology 'Solutionism': a Seller's Market. *Postdigit Sci Educ* 2, 863–878. https://doi-org.inshs.bib.cnrs.fr/10.1007/s42438-020-00164-x

Decoloniality as a Lens of Ethical Foresight for Learning Analytics

Shamya Karumbaiah
University of Pennsylvania
shamya@upenn.edu

Jamiella Brooks
University of Pennsylvania
brooksdj@upenn.edu

Keywords: Educational AI, Ethics, Decoloniality, Colonialism, Philosophy of Learning Analytics

1 EXTENDED ABSTRACT

As Learning Analytics (LA) applications become more widely implemented, the field increasingly engages in the reflective evaluation of the impact of these sociotechnical systems on society. In a recent article, Selwyn (2019) questioned the problematic values and assumptions implicit in LA technologies, and raised criticisms over: 1) blind faith in data, 2) exploitative conditions of the data economy, 3) limits of individual agency, and 4) techno-solutionism. Our field is approaching the inflection point wherein the significant risk to harm already vulnerable peoples will soon outweigh the initial benefits of these technological innovations. As such, we must develop a critical ethical foresight on the intricate connection between our present choices and their implications in the future (Mohamed et al., 2020). The ongoing discourse on fairness and equity in LA needs to be strengthened by contextualizing these inequities in historical systems of power (Sokoloff and Pincus, 2008).

"Coloniality and decoloniality refer to the logic, metaphysics, ontology, and matrix of power created by the massive processes of colonization and decolonization" (Maldonado-Torres, 2016). Furthermore, they seek to explain continuing patterns of power between those advantaged and disadvantaged by colonialism. By applying decoloniality as a methodology, we seek to historicize and humanize individuals, communities, and people groups involved in and affected by LA, as well as to redress harms that have arisen from (and continue to arise from) postcolonial contexts. If "decolonization is a historical process," (Fanon, 1963), then decoloniality seeks to unearth continued colonial processes. With this concept in mind, we aim to leverage Mohamed and colleagues' (2020) taxonomy of decolonial foresight for artificial intelligence to begin examining the coloniality of power both in LA applications (predictions, dashboards) and the structures that support them (e.g., data, policies):

1) *Algorithmic oppression* wherein one social group is privileged over another through data-driven, automated predictive systems (Nobel, 2018). Consider the teacher scoring system developed by Mathematica Policy Research that led to the unfair termination of 206 school teachers in Washington, D.C, United States. In addition to the serious technical flaws, we must also acknowledge that this algorithmic decision making primarily impacted the lives of teachers and students in poor schools, while rich schools continued to invest in people over machines to make such sensitive decisions (O'Neil, 2016).

2) *Algorithmic exploitation* wherein industries asymmetrically benefit from extracting and commodifying education data. For instance, the coerced labor (Selwyn, 2019) of data generation by those students and teachers who do not have autonomy over the commercial LA system used in their classroom or the future use of their data recorded on it.

3) Algorithmic dispossession that leads to the centralization of power in the hands of a few and the disempowerment of the majority (Thatcher et al., 2016). As we continue to engage in the discourse surrounding ethics in LA, we must ask critical questions on who is not included in the conversation, whose interests and concerns are at the center and whose are sidelined, and what values are shaping these conversations.

We argue that a critical foresight is necessary to identify how LA can entrench historical injustices through the continued coloniality of power (Quijano, 2000). Decolonial theories as a philosophical foundation of LA could help us continually examine methodological deficiencies and future harms by 1) connecting algorithmic oppression to the broader social, cultural, and political contexts of learning, 2) raising questions of accountability and responsibility against algorithmic exploitation in educational settings, and 3) interrogating power imbalances in ethics discourse surrounding LA research, design, development, and deployment. Decoloniality offers both temporal breadth and depth necessary to explore these philosophical concerns: it helps us nuance our historical hindsight so that we are learning from the past; it challenges us to reconsider how our present has been informed by that past; and it offers a lens to develop foresight to identify unacknowledged systems of values that shape ongoing advances in our research and development.

REFERENCES

Fanon, F. (1963) *The Wretched of the Earth, trans.* Richard Philcox. New York: Grove, 2004.

Maldonado-Torres, N. (2007) On the coloniality of being: Contributions to the development of a concept. *Cultural studies 21*(2-3):240–270

Mohamed, S., Png, M. T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. Philosophy & Technology, 33(4), 659-684.

Noble, S. U. (2018) Algorithms of oppression: How search engines reinforce racism. NYU Press

O'Neil, C. (2016) *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books

Quijano, A. (2000) Coloniality of power and Eurocentrism in Latin Americas. *International Sociology* 15(2):215–232

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, & S., Vertesi. J. (2019) Fairness and abstraction in sociotechnical systems. In: *Proceedings of the Conference on Fairness, Accountability, and Transparency*, ACM, pp 59–68

Selwyn, N. (2019). What's the problem with learning analytics?. *Journal of Learning Analytics*, *6*(3), 11-19.

Sokoloff, N. J. & Pincus, F. L. (2008) Introduction: Race, class, gender, and capitalism. *Race, Gender & Class* pp 4–8

Thatcher, J., O'Sullivan, D., & Mahmoudi, D. (2016) Data colonialism through accumulation by dispossession: New metaphors for daily data. Environment and Planning D: Society and *Space 34*(6):990–1006

Optimizing Philosophies for Predictive Models in Learning Analytics

Stephen Hutt
University of Pennsylvania
hutts@upenn.edu

Shamya Karumbaiah
University of Pennsylvania
shamya@upenn.edu

Jaclyn Ocumpaugh
University of Pennsylvania
ojaclyn@upenn.edu

Keywords: Machine Learning, Predictive Models, Optimization, Interpretability

As the LAK community begins to synthesize perspectives and priorities from diverse research cultures (Siemens, 2013), one tension is the philosophies used to optimize predictive models. Specifically, some research cultures optimize for *performance* (a common Computer Science/Machine Learning approach), while others optimize for *interpretability* (more common with researchers and practitioners in education). Both philosophies have advantages and are not necessarily contradictory (discussed more below). However, as the learning analytics (LA) community seeks to support learners, the juxtaposition of these philosophies can create challenges. Therefore, we propose that any philosophy of LA should carefully balance these two criteria depending upon relevant contextual factors.

Historically, the machine learning philosophy would advocate for models to be optimized by some measure of performance (e.g., R2, AUROC, F1), often leading to complex and nonlinear interactions. This is clearly advantageous for many contexts, especially high stakes decisions or decisions where the outcome is critical. In such an approach, researchers would select both their model (e.g., Neural Network) and the complexity of that model (e.g., number of layers) in order to train the 'best' model. In contexts like these, it does not matter whether or not the model's inner workings were so intricate that they could be trivially distilled (Chakraborty et al., 2018) (e.g., to a paragraph or page); the complexity does not interfere with the model's use.

In contrast, some educators and researchers take the philosophy that *why* the model is accurate is both more interesting and more relevant than raw performance (Pardo et al., 2016). For example, when predicting learning gains, an intricate, black-box model gives little to no actionable insight regarding how an educational process could be altered or improved without a series of complex simulations and mathematical examination (Chakraborty et al., 2018). The need for meaningful interpretation fundamentally changes the model selection and training philosophy. Researchers may deprioritize performance that comes at the cost of explainability since the latter enables further applications and analysis (e.g., developing interventions to improve the learning experience).

As the LA community merges these two philosophies, the tension between model performance and interpretability can manifest as a real trade-off. Occam's razor (i.e., choosing the simplest model) may not work in this case, where a more complex, less interpretable model might perform significantly better than a more interpretable one. This raises ethical dilemmas that haven't yet been explored well by the LA community. For example, what are the affordances of a more interpretable model over a better performing model and vice versa? To what extent do (or should) model performance be deprioritized to achieve better interpretability? When might it be in students' best interest to deliberately choose a less accurate model? If model interpretability is key, what place do black-box models (e.g., a deep neural network) have in LA? Do these decisions vary based on the learning context?

Many of these challenges we face are also pertinent to conversations happening in the broader modeling/ML community, including those related to algorithmic fairness (Karumbaiah et al., 2019). For example, at what point is the difference in a model's performance for two subpopulations unacceptable? Under what conditions is it ethically sound to reduce the performance for one subgroup in order to boost another? If we cannot truly interpret our model, can we confidently make these decisions?

LA is strengthened by its transdisciplinary nature. However, as it matures, it is vital that we find our philosophy towards optimizing models, likely drawing upon some of the two philosophies discussed here, as well as the contextual issues specific to LA. By merging multiple disciplines' views, we will hopefully gain a deeper understanding and a stronger philosophy moving forward.

REFERENCES

Chakraborty, S., Tomsett, R., Raghavendra, R., Harborne, D., Alzantot, M., Cerutti, F., Srivastava, M., Preece, A., Julier, S., Rao, R. M., Kelley, T. D., Braines, D., Sensoy, M., Willis, C. J., & Gurram, P. (2018). Interpretability of deep learning models: A survey of results. *2017 IEEE SmartWorld Ubiquitous Intelligence and Computing, Advanced and Trusted Computed, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People and Smart City Innovation*. https://doi.org/10.1109/UIC-ATC.2017.8397411

Karumbaiah, S., Ocumpaugh, J., & Baker, R. S. (2019). The influence of school demographics on the relationship between students' help-seeking behavior and performance and motivational measures. *EDM 2019 - Proceedings of the 12th International Conference on Educational Data Mining*, 99–108.

Pardo, A., Jovanovic, J., Mirriahi, N., Dawson, S., Martinez-Maldonado, R., & Gaševic, D. (2016). Generating actionable predictive models of academic performance. *ACM International Conference Proceeding Series*. https://doi.org/10.1145/2883851.2883870

Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*. https://doi.org/10.1177/0002764213498851

Engineering Learning Analytics Technology Environments (ELATE): Understanding iteration between data and theory, and design and deployment

Heeryung Choi
University of Michigan
heeryung@umich.edu

Christopher Brooks
University of Michigan
brooksch@umich.edu

Caitlin Hayward
University of Michigan
cholma@umich.edu

Kirsty Kitto
University of Technology Sydney
Kirsty.Kitto@uts.edu.au

Dragan Gasevic
Monash University
dragan.gasevic@monash.edu

Abelardo Pardo
University of South Australia
abelardo.pardo@unisa.edu.au

Phil Winne
Simon Fraser University
winne@sfu.ca

Neil Heffernan
Worcester Polytechnic Institute
nth@wpi.edu

ABSTRACT: Learning analytics (LA) researchers often fail to collect the data that they need to answer their research questions. The various reasons for the issue range from incorrect tool designs to a lack of understanding on which data is necessary and how educational contexts might affect their data collection protocols. While previous work in other fields suggests that this issue can be addressed through a more cohesive iterative process between the design of the learning experience and the design of the data collection, such a methodology has failed to gain substantial recognition in the learning analytics community. This proposed half-day open workshop aims to enhance the LA community's understanding of how iterative data collection, coupled with theoretically grounded data analysis, can improve the data collection process and the overall quality of a research process. Through the panel talks given by organizers with experiences on data collection tool designs and discussions, participants will share their own experiences of data collection and sketch out potential framework design of the iterative process for their research contexts.

Keywords: Iterative design process, data collection, data analysis, useful data

1 BACKGROUND

The purpose of this workshop is to explore the bidirectional influence of data on theory, and deployment on design, in the construction of educational technologies through the lens of learning analytics. It is often the case that, after the deployment of a novel educational technology, researchers or developers find that their data collection mechanisms do not result in ideal data being captured (Kitto et al., 2020). Possible reasons are a combination of incorrect approaches and the inevitable results of the trial-and-error process; researchers or developers (a) might decide which data to collect based on the functionality of the tool itself, (b) might not fully specify which data to collect due to a lack of understanding of data or tool, or (c) might not consider constraints or characteristics of domain and context where data collection would happen (e.g., time, cost, tools available) (Mostow, 1985). Therefore, after the initial collection of data and conducting exploratory data analysis, researchers often find that they need different granularity levels or additional types of data in order to better map to learner actions, intentions, or cognitive states. This then results in the initial data collection plan needing to be iterated upon, often resulting in adjustments to both what to collect and what tools to use for data collection.

While such an iterative design process may be unavoidable for better data collection and analysis (Jonassen, 2008), there have been few discussions on the processes with respect to both educational software design and data engineering, or what role developers play in improving the data capture of deployed educational technologies. Simply repeating the design process does not guarantee the improvement of data collection (Jonassen, 2008). Successful design requires that researchers and developers, as well as other stakeholders, work together to identify unique data collection contexts and approaches in order to build a rationale behind new designs, a point explicitly addressed by the recently released SoLAR position paper (Kitto et al., 2020). This workshop will seek to cast more light on these engineering tasks, which take into account both the evidence and contexts of deployments.

We propose to develop an approach emphasizing an Evidence-Based Iteration (EBI) method with consideration of contexts. The EBI method gives researchers a framework by which to adjust instrument design, support the generation of more substantive theory, align educational tool design with the learning events that relate to foundational theories, and nurture the sustainable research culture of the educational technology community. The intent of developing this method is not only to better understand the impact of changes made during each round of iteration (e.g. confirmation of theoretical constructs or measurement of an intervention changes), but also to fit data collection needs to situational findings (which will help to move towards theoretical saturation in grounded theory). Furthermore, by encouraging researchers to adopt the EBI method, they will learn the importance of reporting what instrumental features seemingly caused differences, which will benefit both future iterations of the research and future research conducted by the educational technology community more broadly.

The half-day open workshop will bring together people who have experienced the challenges of data collection or developing tools designed to collect data in educational contexts, to share their experiences and to discuss how to improve practices via leveraging the EBI approach.

2 PRE-WORKSHOP PLANS

Participants can submit ½ to 1 page write-up on their previous or current struggle with data collection in particularly when they did not apply a well-planned iterative process. Through this preparatory step, participants can share their context, background, and experience to organizers.

3 WORKSHOP STRUCTURE

Organizers are a diverse group of scholar-practitioners in the learning analytics community, and there will be at least 5 different systems including a learning management system (LMS) for the gameful courses and an online learning platform providing feedback for learners and live report for instructors – and challenges in building these systems – described during the workshop. After an introductory presentation on what iterative process is and why we need it, 6 organizers will start a 60 minutes long panel talk. During the panel talk, they will share (a) the design processes and deployments they went through with the tools, (b) the data they capture, and how it might have changed over iterative development/deployment, and (c) potential theoretical lenses which they have associated with their tools and how those lenses have changed through the design and deployment process. All participants will be encouraged to participate in the panel via questions addressed to the organizers.

The second section of the workshop will be a 30 minute discussion between participants, where they will form a group of 4 to 5 around specific data collection challenges chosen from the responses to the survey. Participants will (a) share participants' own struggle or how they managed to overcome such struggle during data collection or data analysis and (b) find similar and different components of these experiences. The goal of this discussion is to understand similarities and differences in approaches and what caused those and to identify the common top three challenges of their experiences. This discussion section will be followed by 15 minute break.

The third section of the workshop will be another 30 minute group discussion between participants. The goal of this second discussion is to sketch out frameworks (e.g., iterative research workflow) of an iterative process which could work as a potential solution. Participants will be encouraged to think about following questions during the sketch-out: What is identified as a key issue causing the failure or difficulty of ideal data collection? What is the context of the data collection and data analysis they would like to focus on? How can an iterative design process solve the issue? How would researchers know if the designed framework is working? The workshop will be concluded with participants' presentation of their framework design. Since the workshop is virtually offered, there will not be any required equipment for the workshop.

4 OUTCOME AND POST-WORKSHOP PLANS

Anticipated outcomes of the proposed workshop are a white paper that describes (a) why an iterative process is necessary for the educational research, in particularly for data collection and tool design (based on the panel talk and the discussion section 1), (b) how to implement an iterative process, and

(c) how the iterative process can push forward research in the LA community (based on the discussion section 2). The workshop organizers are planning to publicly share the white paper on the SoLAR website or Open Science Foundation website (http://OSF.io). Furthermore, the white paper will be considered as the exploratory data source of the further research on struggles of researchers on data collections and how the EBI method could address the issues.

5 PARTICIPANT RECRUITMENT

Organizers will reach out to various practitioners and software developers to see if they are interested in participating in the workshop. Furthermore, participants will be recruited through the call for participation below:

Learning analytics (LA) researchers often fail to collect the data that they need to answer their research questions due to a lack of in-depth understanding of what types of data they need. While previous works in other fields suggest an iterative design to address the issue, there has barely been a discussion on how to apply iterations on educational contexts in the LA community. This proposed half-day open workshop aims to advance understanding of the necessities of an iterative process in data collection and data analysis based on theories. The workshop includes panel talks by the organizers, who have experience in data collection tool design, and discussion sections to sketch out designs of potential Evidence-Based Iterative (EBI) processes. The EBI process will be able to help LA researchers, practitioners, and other stakeholders update the design of the data collection system in consideration of research contexts and findings from previous iterations. Participants should apply to the workshop by submitting ½ to 1 page write-up on their previous struggle of data collection to the given webpage. Submission should be made by February 9th, 2021.

REFERENCES

Jonassen, D. H. (2008). Instructional Design as Design Problem Solving: An Iterative Process. Educational Technology Research and Development: ETR & D, 48(3), 21–26.

Kitto, K., Whitmer, J., Silvers, A., Webb, M., (2020). Creating data for learning analytics ecosystems. SoLAR position paper. Society for Learning Analytics (SoLAR). Available at: https://www.solaresearch.org/publications/position-papers/

Mostow, J. (1985). Toward better models of the design process. AI Magazine, 6(1), 44–44.

Responsible Learning Analytics: Creating just, ethical, and caring LA systems

Teresa Cerratto Pargman
Stockholm University, Sweden
tessy@dsv.su.se

Cormac McGrath
Stockholm University, Sweden
cormac.mcgrath@edu.su.se

Olga Viberg
KTH Royal Institute of Technology, Sweden
oviberg@kth.se

Kirsty Kitto
University of Technology Sidney, Australia
Kirsty.Kitto@uts.edu.au

Simon Knight
University of Technology Sidney, Australia
sjgknight@gmail.com

Rebecca Ferguson
The Open University, United Kingdom
rebecca.ferguson@open.ac.uk

ABSTRACT: Ethical considerations and the values embedded in the design, development, deployment, and use of Learning Analytics (LA) systems have received considerable attention in recent years. Ethical frameworks, design guidelines, principles, checklists, and a code of practice have contributed a conceptual basis for focused discussions on ethics in LA. However, relatively little is known about how these different conceptual understandings of ethics work in practice and what specific tensions practitioners (e.g., administrators, developers, researchers, teachers, learners) experience when designing, deploying, or using LA with care. This half-day interactive workshop aims to provide participants with a space for information, dialogue, and collaboration around Responsible LA. The workshop will begin with a brief overview of Responsible LA. After that, the participants will present their cases drawing attention to the ethical considerations covered and not covered in LA practices. Following this, participants in groups will discuss the cases illustrating ethical tensions and create semantic categories to document such edge cases. The collected edge cases will be shared in a wiki or database. The workshop outcomes will help inform LA practitioners on ethical tensions that need to be discussed with care while highlighting places where more research work is required.

Keywords: Ethics, fairness, equity, socio-technical systems, values, matters of care, responsible learning analytics

1 BACKGROUND

From its very beginnings, Learning Analytics (LA) has sought to understand the risks associated with a heavy reliance on data and analytics without engaging with the underlying models, algorithms, and assumptions about how students learn (Siemens, 2013). More recently, concerns have arisen in connection with issues regarding potential inequalities (West et., 2016), discrimination (Jones, 2019), data-surveillance (Selwyn, 2019), algorithmic fairness & bias (Holstein & Doroudi, 2019), as well as advisors' rejection of LA systems because of moral discomfort and violation to a professional, ethical code (Jones, 2019). Cases of misuse of students' data have also been reported regarding teachers' lack of data literacy (Lawson et., 2016). These issues are all the more pressing in light of protests from teacher unions in the UK against plans to transform teaching and privatize education data with AI technologies and predictive analytics (Pearson, 2019). Societally, international events have also sparked reflections regarding how structural racism manifests in LA and datasets (Buckingham Shum, 2020). Both research and societal issues concerning data-driven practices in education underscore the seriousness and scope of ethical considerations in the LA community (Cerratto Pargman & McGrath, 2021).

While the LA research community has long been interested in the ethics of data-driven practices (e.g., Slade & Prinsloo, 2013; Pardo & Siemens, 2014; Swenson, 2014; Tsai et al., 2019; Drachsler & Greller, 2016; Sclater, 2016; Ferguson, 2019), most of this work has been conducted in conceptual terms (Arnold & Sclater, 2017). Research on applied ethics has not become pervasive in LA practice, potentially leading to "LA principles and codes of practice being crafted in a theoretical vacuum, far from the practicalities of implementation" (Arnold et al., 2020, p. 2). On this line of reasoning, Kitto & Knight (2019) also stressed the need to engage with concrete cases of the ethics of LA systems "to nurture practical reasoning across the community" (p. 2864).

1.1 Motivation

Capitalizing on past LAK workshops on ethical concerns (e.g., Ferguson et al., 2016; Holstein & Doroudi, 2019; Arnold et al., 2020), this workshop has three primary motivations:

1- Introducing and discussing the interplay between *ethics of justice,* applying rules and principles to ensure the fair and equitable treatment of all people, and *ethics of care*, driven by values and concerns and involving tasks that make a living better in interdependence with others (Puig de La Bellacasa, 2011). Neither ethics of justice nor ethics of care, on their own, can "*sufficiently address and accommodate the complexities, intersectionality and multi-dimensional nature of individuals and different relations in different contexts*" (Prinsloo & Slade, 2017, p.115). Instead, we need to approach ethical considerations in the education sector (K-12, high school, higher education) from a dialectic and relational stance between justice and care.

2- Promoting critical views of data in the context of widespread unease in society about the misuse of data and datasets (D'Ignazio & Klein, 2019). In particular, critical understandings of data are critical to promoting data and ethical literacies.

3- Contributing to ongoing conversations on ethical considerations based on practical cases. It is of utmost importance not only that the LA community is producing ethical frameworks, principles, and concepts but also that it is aware of how ethical considerations can be enacted in practice, what ethical areas are challenging and why, and how the LA community can ensure sustained and updated conversations take place, nurturing practical reasoning on ethics across the community (Kitto & Knight, 2019).

1.2. Relevance to the Conference theme

This workshop is well suited for this year's LAK conference, given the theme of promoting discussions on the impact we make and how we contribute to improved learning. As ethics are deeply entrenched in the learning we scaffold and the teaching we practice, this workshop will encourage the community to reflect on ethical considerations concerning the educational values (i.e., caring for the other) promoted in the design and use of LA systems.

2 WORKSHOP OBJECTIVES AND INTENDED OUTCOMES

The primary goals of this workshop are as follows:

1- Introducing "Responsible LA" via concepts and sensitivities coming from the fields of Science & Technology Studies (Puig de La Bellacasa, 2011) and Human-computer interaction (Buckingham Shum et al., 2019; D'Ignazio & Klein, 2019). By "Responsible LA", we refer to the need to create LA systems that are just and ethical but also that care about equity, democratic and solidarity values in education.

2- Promoting discussions on the ethics of data-driven practices from the ground aimed to inform practitioners on the ethical challenges that emerge in practice.

3- Creating a wiki or other type of artifact contributing to a repository of ethical practice, as suggested by Kiitto and Knight (2019).

4- Helping participants to reflect on ethical challenges that speak of a disconnect between research and practice and find research collaboration opportunities.

The workshop outcomes will advance the LA field by informing the community on ethical challenges encountered in practice (during the development, design, and/or use of LA systems). One concrete outcome of the workshop will be starting an artifact (e.g., wiki) to document edge ethical cases in general terms, not linked to particular individuals or institutions that will be shared in the community for reflective discussions and further study. The workshop outcomes will be disseminated via social media (#ResponsibleLAK) and via the workshop's website:https://sites.google.com/dsv.su.se/responsible-la/home?authuser=0

3 WORKSHOP ORGANIZATION

Type of event: Half-day virtual workshop. Type of participation: Mixed-participation.

The workshop welcomes two participant groups: (1) Those who submit position papers discussing ethical dilemmas they have encountered in their practice (following an open call). These position papers should (a) discuss the context of the case, (b) the ethical concerns, targeting the various stakeholders involved and the principles in tension, and (c) technical, policy, and other approaches that have informed addressing the dilemma, and the effectiveness of these. (2) Those who are interested in attending and participating in the discussions. Submissions will be collated on the workshop website. Publication of the workshop contributions is intended in a joint "LAK Companion Proceedings" (ca. 10 contributions). Participants will post-workshop be invited to contribute to a special issue or similar on "Responsible LA". We expect around 15-20 workshop participants. We will recruit participants via ACM mailing lists, social media, and professional networks.

3.1 Schedule

A preliminary keynote by Rebecca Fergusson (30 min) will be followed by lightning talks in which workshop participants present ethical dilemmas that they have encountered in LA (30 min). The workshop will then move to an interactive mode, where participants break out into smaller discussion centered around selected ethical challenges, working through a series of exercises designed by the organizers to encourage deep thought about the dilemmas they are exploring, the stakeholders impacted, and the ways we might navigate the ethical dimensions of these scenarios (100 min). Finally, the workshop will reconvene to a main plenary, where groups discuss their findings and design the next steps for future work.

3.2 Organizers

We are a group of international scholars with previous experience as workshop organizers in major conferences such as LAK, CSCL, and TEL. We are all from institutions that are investing in Responsible LA and data-driven practices.

4 REFERENCES

Arnold, K., & Sclater, N. (2017). Student perceptions of their privacy in learning analytics applications. https://doi.org/10.1145/3027385.3027392

Arnold, K., Folkestad, J., Ham, M., Pappas, R., Rehrey, G., Shepard, L. (2020). Learning Analytics Principles of Use: Making Ethics Actionable. International LAK Workshop https://lak20laprinciples.weebly.com/

Buckingham Shum, S., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-centred learning analytics. *Journal of Learning Analytics*, *6*(2), 1-9.

Buckingham Shum, S. (2020). Should predictive models of student outcome be "color blind"? http://simon.buckinghamshum.net/2020/07/should-predictive-models-of-student-outcome-be-colour-blind/

Cerratto Pargman, T., & McGrath, C. (2021). Mapping the Ethics of Learning Analytics in Higher Education: A Systematic Literature Review of Empirical Research. Journal of Learning Analytics, 1-17. https://doi.org/10.18608/jla.2021.1

Drachsler, H., & Greller, W. (2016). Privacy and analytics – it's a DELICATE issue. A checklist for trusted learning analytics. https://doi.org/10.1145/2883851.2883893

D'Ignazio, C., & Klein, L. (2020). *Data feminism*. MIT Press.

Ferguson, R., Hoel, T., Scheffel, M., & Drachsler, H. (2016). Guest editorial: Ethics and privacy in learning analytics. *Journal of Learning Analytics, 3*(1), 5-15.

Ferguson, R. (2019). Ethical challenges for learning analytics. *Journal of Learning Analytics, 6*(3), 25-30.

Holstein, K & Doroudi, S. (2019). International LAK Workshop on Fairness and Equity in Learning Analytics Systems (FairLAK). https://sites.google.com/view/fairlak

Jones, K. (2019). Advising the whole student: eAdvising analytics and the contextual suppression of advisor values. *Education and Information Technologies, 24*(1), 437–458.

Kitto, K., & Knight, S. (2019). Practical ethics for building learning analytics. *British Journal of Educational Technology, 50*(6), 2855-2870.

Lawson, C., Beer, C., Rossi, D., Moore, T., & Fleming, J. (2016). Identification of 'at risk' students using learning analytics: the ethical dilemmas of intervention strategies in a higher education institution. *Educational Technology Research and Development, 64*(5), 957–968.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. British Journal of Educational Technology, 45(3), 438–450.

Pearson (2019). Pearson 2025. Transforming teaching and privatizing education data. Retrieved 2010/03/17.

Prinsloo, P., & Slade, S. (2017). Big data, higher education and learning analytics: Beyond justice, towards an ethics of care. In Big Data and Learning Analytics in Higher Education (pp. 109–124). Springer.

Puig de La Bellacasa, M. P. (2011). Matters of care in technoscience: Assembling neglected things. *Social Studies of Science, 41*(1), 85–106.

Sclater, N. (2016). Developing a code of practice for learning analytics. *Journal of Learning Analytics, 3*(1), 16–42.

Selwyn, N. (2019). What's the Problem with Learning Analytics? *Journal of Learning Analytics*, *6*(3), 11-19.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10, SI), 1380–1400. https://doi.org/10.1177/0002764213498851

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1510–1529. https://doi.org/10.1177/0002764213479366

Swenson, J. (2014). Establishing an ethical literacy for learning analytics. https://doi.org/10.1145/2567574.2567613

Tsai, Y., Poquet, O., Gašević, D., Dawson, S., & Pardo, A. (2019). Complexity leadership in learning analytics: Drivers, challenges and opportunities. *British Journal of Educational Technology, 50*(6), 2839–2854. https://doi.org/10.1111/bjet.12846.

West, D., Huijser, H., & Heath, D. (2016). Putting an ethical lens on learning analytics. Educational Technology Research and Development, 64(5), 903–922. https://doi.org/10.1007/s11423-016-9464-3

Privacy-by-Design for Responsible and Equitable LA Systems, Policies, and Practices

Carrie Klein
George Mason University/Future of Privacy Forum
cklein7@gmu.edu

ABSTRACT: Learning analytics (LA) systems are increasingly being employed by postsecondary institutions to support student outcomes. Institutional capacity to collect vast amounts of data about students and their environments means that LA systems are also increasingly being used beyond classroom and advising sessions to support institutional outcomes. In this position paper, I use a case study of two United States (U.S.) research universities to show how LA systems are used in practice by administrators seeking to advance both institutional and student success. I illustrate that while LA systems data in concert with other institutional data work can advance institutional goals and support individual students, use of these systems has also resulted in increased surveillance of students and inequitable student outcomes, particularly for students from marginalized populations. I argue that the conflation of institutional and student success efforts works against the goals of just, ethical, and caring LA systems use. To improve responsible use of LA systems needed is a Privacy-by-Design approach that embeds privacy, upstream, into systems, policies, and practices as a lever to mitigate power imbalances and leverage equitable data use.

Keywords: Learning analytics; surveillance; equity; privacy-by-design; power.

1 OVERVIEW: THE NEED FOR AN UPSTREAM, PRIVACY-BY-DESIGN APPROACH TO RESPONSIBLE LA SYSTEMS

Postsecondary institutions are employing learning analytics (i.e., big data about students and their contexts) to support student learning and improve student outcomes (Siemens, 2013). In an era of increased accountability, institutions are being called on to show evidence of their effectiveness (e.g., retention, progression, and completions metrics; Kelchen, 2018; Prinsloo & Slade, 2014). As such, student data, like those in LA systems, are an increasingly valuable resource for institutions, which are under pressure to prove both their students' and their own success. Using a collective case study of two U.S. research universities, I illustrate how the conflation of goals and of data has resulted in inequitable uses of LA systems and increased surveillance of students.

Through case study evidence, I argue that while institutions may adhere to the tenets of responsible data governance and use and be committed to equitable student outcomes, the inherent value of student data to institutional success exacerbates the power differential between institution and individual. This muddies organizational determinations of ethical, just, or care-based uses of data and justifies expansive data collection of student data, increased surveillance and inequitable outcomes. To remedy this, requires an acknowledgement of the role of power between the institution and the individual and how the various context in which LA systems are used can reaffirm, rather than remediate structural inequities. Also needed is an upstream approach to integrating ethics, justice and care. Ethics historically were viewed as an afterthought, or even barrier, for LA and other socio-technical systems (Ferguson, et al., 2016). Data justice and care, while vital for centering data subjects in processes, are often bestowed by the institution or the state onto the individual through various rights and responsibilities (Prinsloo & Slade, 2017; Taylor, 2017). Using a privacy-by-design approach that acknowledges power dynamics, integrates ethics, justice, and care, and centers center student agency may foster more equitable and, ultimately, responsible use of LA systems from inception through application.

2 CASE CONTEXT: CONFLATION OF STUDENT AND INSTITUTIONAL SUCCESS EFFORTS LEADS TO INCREASED SURVEILLANCE AND INEQUITABLE OUTCOMES

2.1 Case Details

As a part of a larger collective case study to investigate the institutional logics (e.g., assumptions, values, norms, and 'rules') of LA systems and other student data, I interviewed 55 participants at two public research universities in the southeastern region of the U.S. Each of the universities was actively using student analytics data to support student success efforts and improve institutional outcomes. One university, Flagship Public University (FPU) is the oldest public university in the state, highly selective, and composed of a predominantly White and highly academically prepared student population. The other university, Access Public University (APU) is one of the newest research institutions in the state, is access-oriented, and is composed of a predominantly Black student population with variable preparedness for postsecondary academic work. Participants, who made up the bulk of the interviews, held positions across administrative and teaching ranks, to include vice presidents, assistant provosts, information technologists, institutional effectiveness researchers, faculty, student affairs professionals, and advisors.

2.2 Findings

Both institutions were using LA systems and other student data to improve student outcomes. Notably, none of the participants differentiated LA data from other forms of student data – it was all just data in their minds. At both FPU and APU, participants articulated that they used LA systems data to improve decision making and support student success. However, how they conceptualized success differed. At FPU, successful support of students through LA systems meant mitigating mental health considerations, maximizing educational co-curricular experiences (e.g., study abroad opportunities or internships) and opportunities for academic exploration, and increasing selectivity as a means to improve institutional metrics. At APU, student success was focused squarely on credentialing as a means to improve institutional metrics and individual outcomes, meaning graduating more students in less time and with less debt. This often meant encouraging students toward specific educational pathways and limiting academic exploration. Although participants at both institutions expressed a deeply held commitment to equity, the different contexts associated with these institution results in decidedly different, and arguably inequitable educational experiences and educational opportunities.

The drive by both FPU and APU to meet not just mission-oriented student support goals, but also institutional accountability goals also resulted in ramping up of collection of student data and student surveillance. At APU, institutional technologists noted that they work working toward having a "digital footprint" of every measurable student interaction, from the moment they set foot on or logged on to campus. While FPU was less aggressive with their surveillance measures, they were still gathering vast amounts of data about their students. Notably, although both institutions were concerned about appropriate and responsible use of LA systems and other student data, at the time of my research, neither had put into place data governance plans that spoke specifically to issues of ethics, justice, or care or worked to limit data collection or protect student data privacy (beyond basic FERPA compliance). While an APU vice president noted that students at the institution were informed about the data that were collected about them, they were not necessarily aware of how those data were used to inform institutional decision making, beyond advising and financial aid support. FPU took a different approach, by not sharing with students the extent of data that are collected about them or how that data is used. In fact, an FPU assistant provost noted that, as an institution, they were using student data to inform decision making but had not yet addressed issues of ethical or equitable use of LA systems and other student data. The disconnect between employing data, culled from students and their experiences, without informing the students who are providing institutions with this valuable resource (i.e., their information as data subjects) or acknowledging, explicitly, how that data will be collected, used, and protected, speaks to a disconnect between ethics and practice and to an uneven balance of power between the institution and the data subject.

3 ETHICAL CONCERNS: SURVEILLANCE HARMS AND PRIVACY AS POWER

3.1 Surveillance Harms

The increased data collection and surveillance of students at APU and FPU is tied to the nature of socio-technical systems. Big data and algorithms, like those associated with LA systems are "data hungry" (Obermeyer & Emanuel, 2016). This hunger in concert with the potential for LA systems to benefit institutional decision-making acts as a justification for collecting more data through increased student surveillance. Zuboff (2019) and Monahan (2008) have argued that surveillance systems within capitalist systems, like those associated with big data technologies in western education, run counter to democratic principles by exploiting individual information for institutional gain. The tension between the nature of LA systems and the mission of higher education within the context of an increasingly corporatized and differentiated postsecondary landscape creates foundational ethical concerns for the use of LA systems to assist students – even when that use is well-intentioned by postsecondary institutions.

Compounding this foundational tension is the propensity for systems of surveillance to inequitably harm marginalized populations. The inequitable surveillance of people of color, people with disabilities and people with lower socio-economic statuses, especially within education has been well documented (Cyphert, 2020; Schroeder, 2016). Gilliard (2019) argues that surveillance is a *de-facto* state of affairs for members of these groups, and Schroeder (2016) argues that inequitable surveillance has resulted in increased discipline and tracking of students of color in education. With the advent of COVID-19, the expansion and harms of inequitable surveillance has been magnified (Cyphert, 2020). Inequitable surveillance leads to can digital redlining (Gilliard & Culik, 2016) by limiting opportunities and outcomes for marginalized populations. The APU and FPU case illustrates how well-intentioned, but conflated, success efforts can result in institutional harm to students by creating inequitable educational opportunities and reifying structural inequities.

3.2 Privacy as Power

In most cases students are minimally aware of and usually uniformed about what data are collected about them by their institutions, how their data is being used, how that use will impact them long-term or who owns their data (Brown & Klein, 2020). Lack of student awareness and agency an example of the inherent information asymmetry that exists within “imbalanced power relationships” (Tsai, et al., 2020, p. 2), like those in LA systems. Véliz (2020) argues that privacy is power, and it can work to rebalance democratic systems of engagement to ensure the responsible use of socio-technical systems. By pairing privacy with ethics, justice and care, postsecondary institutions can remedy the inherent power imbalance between themselves and students and further responsible use of LA systems to benefit both students and their schools.

4 ADDRESSING THE DILEMMA: EMBEDDING PRIVACY, MODERN POLICIES, AND RELEVANT PRACTICES FOR RESPONSIBLE LA DATA USE

As with most innovations, the appropriate policies and practices to govern responsible LA system use has often lagged behind the innovation, itself. For instance, postsecondary data privacy policies do not explicitly address the nature of modern data systems, the inherent power dynamic at play, or the needed equity measures to mediate that dynamic (Brown & Klein, 2020). If we acknowledge that privacy, in relation to surveillance, is tied to power and if we are committed to an ethical use of student data predicated on notions of justice and of care, needed is an upstream, embedded, modern, and relevant approach to promote responsible (i.e., ethical, just, and care-based) use of LA systems and other student data.

4.1 Privacy-by-Design to Upstream and Embed Responsible Data Use

Cavoukian (2006) argues that with the growth of socio-technical systems and a knowledge economy that values data as a resource, privacy “must be incorporated into networked data systems and technologies, by default” (p. 2, para. 4). She provides seven principles for privacy-by-design and argues that by proactively embedding privacy principles into the makeup of systems, policies, and practices from, privacy can be more effectively protected and users more effectively centered in the process. The seven principles included in privacy-by-design are: 1) Proactive not reactive and preventative privacy; 2) Privacy as a default setting; 3) Privacy embedded into design; 4) Positive-sum, not zero-sum privacy; 5) End-to-end/lifetime security; 6) Visibility and transparency; and 7) User-centered (Cavoukian, 2006, p. 6).

While Cavoukian does not explicitly name ethics, justice, or care, I argue that the privacy-by-design principles she espouses sets up a foundation for ethical, just, and care-based policies and practice. Incorporating privacy principles upstream can create improved opportunities for addressing concerns in the development and day-to-day use of LA systems, like those related to surveillance or data use resulting in inequitable outcomes. Of course, the criticism of privacy-by-design is that it is yet another set of principles that may or may not be applied in practice. However, I contend that the value of the privacy-by-design approach is that, if applied regularly and upstream in the development of technology, policy, and practice, it can address issues of over-surveillance, over-collection of data, and inequitable use of data before those issues emerge - paving the way for more responsible and equitable use of LA systems and data.

4.2 Modern Policies and Appropriate Practices

Privacy-by-design can also inform the development of modern data privacy policies. Certainly, there has been some legal work done to modernize data privacy and improve the power dynamic between institutions, data systems, and individuals (e.g., the GDPR in Europe and the right to be forgotten). In the U.S. we have patchwork of state laws, federal FERPA regulations, and various institutional policies, but these policies often do not address modern conceptions and uses of data or the privacy-related power dynamics that exist. In our recent work on data privacy policies in U.S. higher education institutions (Brown & Klein, 2020), we found that data are seen as an institutional resource but are also often conceptualized as static student records– rather than vast data systems, able to be recombined and repurposed for a variety of purposes. We argued that this conception ignores the reality of dynamic socio-technical systems associated with LA and other analytics data used by institutions to inform decision making. Further, data privacy policies are often poorly communicated to students, limiting their usefulness for informing students of their rights or potential inequities or harms related to the collection and use of their data or any associated surveillance or success initiatives (Brown & Klein, 2020).

From a privacy-by-design perspective, effective policies would establish appropriate practices that benefit both institutional and student goals and be student-centred and transparent – with clearly communicated standards for collection, use, retention, destruction, and ownership of modern data systems and opportunities for consent, redress, and opting out of LA systems. In the case of FPU and APU, use of privacy-based policies embedded into practice might mitigate over-collection and over-surveillance of students, minimize inequitable outcomes, and foster opportunities for increased student agency. Policies informed by privacy-by-design must be paired with privacy-informed practice. From a power perspective, institutions, educators, and technologists should consider whether informed practice or the affordance of a technology, like LA systems are driving use of data. If LA systems-related practices are transparent, proactive, and preventative in nature, they are more likely to minimize opportunities for digital redlining, bias creep, and over collection of data and align with ethical, just, and care-based priorities for responsible use.

REFERENCES

Brown, M., & Klein, C. (2020). Whose data? which rights? whose power? A policy discourse analysis of student privacy policy documents. *The Journal of Higher Education*, *91*(7), 1149-1178.

Cavoukian, A. (2006, Nov.). Creation of a global privacy standard. Toronto, Ontario: Canada Information and Privacy Commissioner of Ontario. Retrieved from: https://iapp.org/media/pdf/resource_center/pbd_implement_7found_principles.pdf

Cyphert, A. (2020). *Tinker*-ing with machine learning: The legality and consequences of online surveillance of students. *Nevada Law Journal 20*(2): 457-501.

Ferguson, R., Hoel, T., Scheffel, M., & Drachsler, H. (2016). Guest editorial: Ethics and privacy in learning analytics. *Journal of learning analytics*, *3*(1), 5-15.

Gilliard, C. (2019, Mar. 25). Privacy's not an abstraction. *Fast Company*. Retrieved from: https://www.fastcompany.com/90323529/privacy-is-not-an-abstraction

Gilliard, C. & Culik, H. (2016, May 24). Digital Redlining, Access, & Privacy. Common Sense Education. Retrieved from: https://www.commonsense.org/education/articles/digital-redlining-access-and-privacy

Kelchen, R. (2018). *Higher education accountability*. Baltimore, MD: Johns Hopkins University Press.

Monahan, T. (2010). Surveillance as governance: social inequality and the pursuit of democratic surveillance. In K.D. Haggerty & M. Samata (Eds.). *Surveillance and democracy*. New York, NY: Routledge.

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. *The New England journal of medicine*, *375*(13), 1216.

Prinsloo, P., & Slade, S. (2014). Student data privacy and institutional accountability in an age of surveillance. In A.M. Menon, et al. (Eds.), *Using data to improve higher education* (pp. 197-214). Rotterdam, The Netherlands: Sense Publishers

Prinsloo, P., & Slade, S. (2017). Big data, higher education and learning analytics: Beyond justice, towards an ethics of care. In *Big data and learning analytics in higher education* (pp. 109-124). Springer, Cham.

Schroeder, S. Surveillance, violence, and the marginalization of students of color. *Perspectives on Urban Education*, *13*(1): 56-59

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57: 1380. DOI: 10.1177/0002764213498851

Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, *4*(2.

Tsai, Y. S., Whitelock-Wainwright, A., & Gašević, D. (2020, March). The privacy paradox and its implications for learning analytics. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 230-239).

Véliz, C. (2020) *Privacy is power: Why and how you should take back control of your data.* New York, NY: Bantam Books.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power.* New York, NY: Public Affairs.

Reframing Student Privacy as a Common Value and Responsibility

Kyle M. L. Jones
Indiana University-Indianapolis (IUPUI)
kmlj@iupui.edu

Tyler Dell
Indiana University-Indianapolis (IUPUI)
tdell@iu.edu

ABSTRACT: In the American higher education context, student privacy is treated as an individual right. In this workshop paper, we argue that in light of emerging sociotechnical conditions this approach is flawed. Data mining, predictive analytics, machine learning, and artificial intelligence continue to push the boundaries of student privacy in ways once unimaginable, all of which challenge federal law, institutional policy, and contextual norms. Instead of relying on existing, non-workable conditions to protect students, we argue that institutional actors need to reframe their thinking about student data and student privacy by taking up a position that the data is a common-pool resource and privacy is a shared value—and responsibility.

Keywords: Higher education, student privacy, common-pool resources

1 LEARNING ANALYTICS AND THE JUSTIFIED ETHICAL BAR

The aggregation and analysis of student data for learning analytics purposes has led to sustained, critical conversations around ethics and policy since the field's genesis in 2010. And unlike as often happens with other sociotechnical developments, the ethical issues quickly moved from a "marginal" to a central position for learning analytics practitioners and researchers (Prinsloo & Slade, 2017b, p. 50). Since learning analytics are predicated on the ability to surface an array of student data, and doing so would clearly challenge established norms, expectations, and policy principles, the ethics of the work had to be addressed. Numerous research projects responded by attempting to address myriad ethical issues and promote policy principles to guide the field (see Ferguson, 2012; Khalil & Ebner, 2015; Pardo & Siemens, 2014; Prinsloo & Slade, 2013; Rubel & Jones, 2016; Sclater, 2016; Slade & Prinsloo, 2013).

If learning analytics was solely concerned with the data-driven study of learning behaviors to improve learning outcomes, then the ethical issues would be fairly tractable. Core learning analytics researchers profess that "learning analytics is about learning" (Gašević et al., 2015), but the fact of the matter is that learning analytics has been co-opted as a cultural and political tool (Prinsloo & Slade, 2017a). The methods, mesures, and infrastructures on which learning analytics rely have enabled administratively-minded actors and the surveillant-curious the ability to datafy the student experience and attempt—with varying degrees of success—to nudge, direct, and manipulate behaviors to serve bureaucratic interests. The end result has been a power-driven rush by administrators to make available for analysis nearly any available data, most often about students but not always, to support their neoliberal philosophy and evidence-based decision-making goals (Cannella & Koro-Ljungberg, 2017; Giroux, 2013). This techno-managerialism has, according to Ball (2004, p. 144, as cited in Prinsloo, 2020), led to a culture of performativity "and a mode of regulation, or even a system of 'terror'..., that employs judgements, comparisons, and displays of means of control, attrition, and change." There are serious concerns that universities, when they use student data as a means to serve institutional and education-adjacent ends, fail to balance their own interests with those of their students (Borgman, 2018; Jones, Rubel, et al., 2020). Learning analytics, then, is much more than just an emerging interdisciplinary educational science with a small core of ethical problems—it is a politically-charged epistemology ripe with ethical conflicts and questionable values its scholarly progenitors would be unlikely to support.

The swamp of ethical problems centers on one key concern: student privacy. Learning analytics falls on its face if data and information about students is not created, mined, and acted upon. Because of this, student privacy protections are seen by some as "obstacles" (Drachsler et al., 2015) to achieving learning analytics goals. Much intellectual effort has been put into unpacking student privacy empirically and philosophically with the aim of coming up with a greater understanding of its facets and some workable governance solutions. We know from published research that the particular student privacy issues are neither narrow nor resolved. For instance, there is an open debate about the extent to which students should have control over data that identifies or concerns them (Reidenberg & Schaub, 2018; Tene & Polonetsky, 2013), and this leads to other questions about data ownership (Regan & Jesse, 2018). Other work has raised interesting, but perhaps not tractable, points about students' rights to be informed and consented about learning analytics (Jones, 2019; Prinsloo & Slade, 2018). Furthermore, the COVID-19 pandemic and the en masse move to online education has heightened sensitivities regarding surveillance and the degree to which edTech companies are financially benefiting from the surfeit of student data higher education is creating as it relies on these technologies, especially proctoring tools (Coghlan et al., 2020; Morrison & Heilweil, 2020). Some higher education actors have—seemingly without concern for student privacy—called for the surveillance of online student behaviors just because students are at the present time completely reliant on edTech to pursue a degree (Tufekci, 2020; Zimmerman, 2020).

On the empirical front, we are gaining a much clearer picture about students' privacy expectations and perceptions vis-à-vis learning analytics. Students have a growing awareness of the data they disclose within learning management systems (Vu et al., 2020), but remain fairly uninformed about other analytic pursuits (Jones, Asher, et al., 2020). Students report understanding the goals and possible benefits of learning analytics (Karaoglan Yilmaz & Yilmaz, 2020; Nevaranta et al., 2020). Still, they hold privacy reservations (Jones, Asher, et al., 2020; West et al., 2020). Some are worried about data management and security practices especially by third parties, stating that data systems are often infiltrated (Nevaranta et al., 2020), but they tend to trust their respective institution's data practices (Jones, Asher, et al., 2020; Vu et al., 2020). The through line within this subset of the empirical literature is that students want to be more informed about learning analytics and express some agency (e.g., limiting access, information control) over personally identifiable data and information (Jones, Asher, et al., 2020; Roberts et al., 2016; Slade et al., 2019; Tsai et al., 2020).

These and many other ethical issues require careful unpacking to address them in research and practice. Such challenging ethical work has Fritz and Whitmer (2019, p. 27) throwing up their proverbial hands with frustration, writing, “increasingly, when we observe the research and practice on learning analytics, it is difficult not to be overwhelmed about the ethical implications of using student data…. [leading] some to reasonably conclude ‘why bother?’” We “bother” with the justified, high ethical bar in front of learning analytics because to not do so would be morally corrupt, and because the foundation for learning analytics would be weak and unstable should we choose not to do so. We reluctantly admit, however, that the state-of-the-art with learning analytics—the methods, the technologies, the policies—leave us questioning if student privacy can and will ever become a protected right or just something at which advocates of educational data mining handwave.

2 DEGRADING CONDITIONS FOR STUDENT PRIVACY

Our current thinking on privacy stems from an observation that the conditions necessary to protect students have not measurably or notably improved since the inception of learning analytics. Ten-plus years have passed, and only minor movements to protect student privacy have been made. Some institutions have learning analytics-specific policies (see Colorado State University, n.d.; University of California-Berkeley, n.d.; University of Hawaiʻi at Mānoa, n.d.), while others have hired chief privacy officers (see Johnson, 2019; Vogel, 2015). There has been some advocacy on behalf of students, especially around issues with facial recognition (Burke, 2020). And major privacy organizations, like the Electronic Frontier Foundation (n.d.) and the Future of Privacy Forum (n.d.), have developed useful informational resources. 17 States have passed some form of privacy law, but no major amendments to the United States federal student privacy law (FERPA) address emerging issues brought about by learning analytics—in fact the opposite is true. The December 2, 2011 amendment arguably enlarged who could gain access to personally identifiable student data and information by naming federal agencies access and researchers as permitted entities (Electronic Privacy Information Center, n.d.; *Family Educational Rights and Privacy; Final Rule, 2011*, 2012).

Some may argue that these changes are enough to suggest that student privacy *is* a right that is being protected now more than ever. But how can that be the case when, for example, instructors and their institutions force students into surveillant infrastructures that, as is the case with Proctorio, record and analyze their every move while causing anxiety in the name of academic integrity (Chin, 2020)? How can that argument stand when institutions have given away geolocation data to campus police for investigatory purposes, as Indiana University did (Liberty Justice Center, 2020)? How does this position hold up when institutions are working as consortia to develop massive student data warehouses for future purposes widely unknown and with policies inaccessible to students (Lederman, D., 2018)? No sweeping privacy-enhancing technologies and practices have been developed, no clear privacy-protecting norms have emerged, and no major legal changes have been made. In the absence of these things, learning analytics practices, methods, and the infrastructures that support them march forward—and in doing so they potentially increase opportunities for unethical data access and use in ways that put students’ privacy and autonomy at risk.

If student privacy-protecting conditions are getting worse, even with awareness that student privacy is threatened by learning analytics, then it follows that our current practical, ethical, and legal approaches are somehow failing. We envision three possible explanatory reasons. First, these things are simply ill-matched in an era of big(ger) data in higher education. Previously, the sociotechnical ecosystem produced student data in smaller quantities, with less granularity, and held less value and only for a few higher education actors. These things have been turned on their head: data is ubiquitous, finely grained, and immensely valued—socially and commercially—by actors internal and external to an institution (Selwyn, 2015). Perhaps the privacy problem is just too vast, too technical, and too socially charged to be addressed with any extended resolve.

Second, a point that is not entirely divorced from the first, the current legal and institutional policy frameworks cannot account for new student privacy needs. Law-making works at a glacial pace typically unfit for our frenetic technological society. And since institutional policy is influenced by existing legal frameworks, there has been little motivation to react to the changing (and very real) challenges to student privacy.

We think a third reason holds the most explanatory power: there is a reason to *not* change policy or practice. Learning analytics advocates have a vested interest in maintaining the status quo: the Wild West of student data flows enables the sociotechnical imaginary they see as a future reality. Why would they push for privacy protections that may throttle down student data, the lifeblood of their work? In the American legal history of privacy, the battle of values—individual rights versus some other end or interest—privacy tends to lose out. About this Daniel Solove (2008) writes:

> The interests aligned against privacy—for example, efficient consumer transactions, free speeches, or security—are often defined in terms of their larger social value. In this way, protecting the privacy of the individual seems extravagant when weighted against interests of society as a whole. (p. 89)

In this light, student privacy is seen as an oppositional force. But against what? Education. The most fervent of learning analytics advocates argue for the transformative power of data mining and analytics to remake educational institutions, to serve as a “pedagogic corrective” (Selwyn, 2013, p. 33) that moves colleges and universities to become more efficient, effective, and personalized. To the advocates, student privacy slows these processes and the aligned goals.

3 REFRAMING STUDENT PRIVACY

It seems that the only way to position student privacy as a key value worth attention and action is to reframe it as something that is an *undeniable, intrinsic* element of education itself—which is what we argue. To suppress student privacy protections would be to limit education’s progress, for the ability to consider intellectual ideas and develop speech free from influence or manipulation is a critical part of learning. In this paper, we *do not* accept the position that higher education is primarily an economic engine or a commercial enterprise. If this were the case, then learning gains would not matter: only cost/benefit ratios, tuition dollars, employment numbers, and the like would. We still believe, like Amy Gutmann (2015) does, that most higher education institutions—excepting for-profits—primarily care about increasing growth opportunities for their students, exposing students to intellectual ideas and the application of those ideas in various domains (what Gutmann calls “creative understanding”), and to help students contribute to their communities and wider society.

One cannot deny the appropriate flow of student data and information to achieve an institution’s analytic ends. No one can neither defend a position that *all* student data and information should stop flowing nor that *all* learning analytics are unethical. A university contains multitudes of informational needs—educational, administrative, and otherwise—that require access to student data because students are the key stakeholders of an institution. It is at this group that institutional efforts are directed. But it is also because of this truth that the responsibility to protect student privacy is distributed. Student privacy is a *common value* and student data is a *common-pool resource*. An advisor, faculty member, and administrator all pull from the pool to achieve their individual information goals, but they also share the responsibility to protect that pool’s resources overall and when in use.

We are not the first to align personal information and data with a common-pool conceptualization. Priscilla Regan (2000) was arguably the first on this privacy scene. Drawing insights from Elinor Ostrom’s economic theory of the tragedy of the commons, Regan conceptualizes cyberspace as a commons and personal information as a common-pool resource. Such resources, according to Ostrom (1990, p. 30), are part of “a natural or man-made resource system that is sufficiently large as to make it costly (but not impossible) to exclude potential beneficiaries from obtaining benefits from its use.” Data and information about students exist within human-made resource systems. A commons has a limit to the amount of beneficiaries and uses which it can support. If this maximum carrying capacity is exceeded, the quality of the commons—and the quality and quantity of its resources—will degrade, unless co-operative use agreements are created and followed. These agreements can be normative, codified in policy, and even embedded in code.

Learning analytics (and other information practices) requires access to the common pool’s resources, and as we wrote above, it would be costly to deny institutional actors to the student data within the pool. What is plausibly more costly than restricted access? Improper and socially harmful uses of the common-pool’s resources. Unlike a natural resource—say a fishery—the quantity of information in the student information pool does not deplete with use since information is a non-rivalrous good. But, like a fishery, there are quality issues with mis- and overuse. For example, unregulated fishing in a fishery often results in a fish shortage and polluted water. Similarly, underregulated use of student data leads to potential analytic outcomes that are biased, developed without rigor, or misleading. These analytics, or even unanalyzed, so-called “raw” data could then be added back to the common pool to problematize downstream uses. These are technical issues arising from poor resource stewardship, but there are also real social harms.

Degradations of the commons, a pool of resources for which nearly all institutional actors hold some responsibility to protect and steward, can directly lead to student claims of privacy harms. The misuse of identifiable and re-identifiable data pushes students into a corner where they begin to distrust their institutions. If this distrust reaches an apogee, student backlash occurs: relationships break, institutional reputations are tarnished, legal remedies are pursued, and—most importantly—the pursuit of one’s education is put at risk.

4 CONCLUDING THOUGHTS

This paper lays out a working framework for thinking about and acting on student privacy using a communitarian lens. Its central thesis is that student data and information is a common-pool resource that loads institutional actors with the responsibility to protect and use the resource with care in order to respect student privacy rights. The reframing seems necessary because the individual rights approach is limited and no existing motivations seem to be moving the student privacy ball forward to get learning analytics advocates and others to take up substantive work on this very important value. The nature of this work—a workshop paper—leaves a number of open questions, some of them practical, others theoretical, and others yet philosophical. And as a work-in-progress, it has rough edges. The authors welcome substantive, constructive comments to help develop a strong foundation for this framework.

REFERENCES

Borgman, C. L. (2018). Open data, grey data, and stewardship: Universities at the privacy frontier. *Berkeley Technology Law Journal*, *33*(2), 365–412. https://doi.org/10.15779/Z38B56D489

Burke, L. (2020, February 21). UCLA drops plan to use facial recognition security surveillance, but other colleges may be using technology. *Inside Higher Ed*. https://www.insidehighered.com/news/2020/02/21/ucla-drops-plan-use-facial-recognition-security-surveillance-other-colleges-may-be

Cannella, G. S., & Koro-Ljungberg, M. (2017). Neoliberalism in higher education: Can we understand? Can we resist and survive? Can we become without neoliberalism? *Cultural Studies ↔ Critical Methodologies*, *17*(3), 155–162. https://doi.org/10.1177/1532708617706117

Chin, M. (2020, April 29). Exam anxiety: How remote test-proctoring is creeping students out. *The Verge*. https://www.theverge.com/2020/4/29/21232777/examity-remote-test-proctoring-online-class-education

Coghlan, S., Miller, T., & Paterson, J. (2020). Good proctor or "Big Brother"? AI ethics and online exam supervision technologies. *ArXiv*. http://arxiv.org/abs/2011.07647

Colorado State University. (n.d.). *Ethical principles of learning analytics*. https://www.chhs.colostate.edu/alt/ethical-principles-of-learning-analytics/

Drachsler, H., Hoel, T., Scheffel, M., Kismihók, G., Berg, A., Ferguson, R., Chen, W., Cooper, A., & Manderveld, J. (2015). Ethical and privacy issues in the application of learning analytics. *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, 390–391. https://doi.org/10.1145/2723576.2723642

Electronic Frontier Foundation. (n.d.). *Student Privacy*. Electronic Frontier Foundation. Retrieved February 5, 2021, from https://www.eff.org/issues/student-privacy

Electronic Privacy Information Center. (n.d.). *Family Educational Rights and Privacy Act (FERPA)*. Retrieved February 9, 2021, from https://epic.org/privacy/student/ferpa/#history

Family Educational Rights and Privacy; Final Rule, 2011, 76 Fed. Reg. 75604 (2012), (to be codified at 34 C.F.R. pt. 99).

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *Journal of Technology Enhanced Learning*. http://inderscience.metapress.com/index/W1QP4L6217K0Q2PV.pdf

Fritz, J., & Whitmer, J. (2019). Ethical learning analytics: "Do no harm" versus "do nothing." *New*

Directions for Institutional Research, *2019*(183), 27–38. https://doi.org/10.1002/ir.20310

Future of Privacy Forum. (n.d.). *The Student Privacy Resource Center*. Student Privacy Compass. Retrieved February 5, 2021, from https://studentprivacycompass.org/

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

Giroux, H. A. (2013). Neoliberalism’s war against teachers in dark times. *Cultural Studies ↔ Critical Methodologies*, *13*(6), 458–468. https://doi.org/10.1177/1532708613503769

Gutmann, A. (2015). What makes a university education worthwhile? In H. Brighouse & M. McPherson (Eds.), *The aims of higher education: Problems of morality and justice* (pp. 7–25). The University of Chicago Press.

Johnson, S. (2019, March 25). Chief privacy officers: A small but growing fleet in higher education. *EdSurge*. https://www.edsurge.com/news/2019-03-25-chief-privacy-officers-a-small-but-growing-fleet-in-higher-education

Jones, K. M. L. (2019). Learning analytics and higher education: A proposed model for establishing informed consent mechanisms to promote student privacy and autonomy. *International Journal of Educational Technology in Higher Education*, *16*(1), 24. https://doi.org/10.1186/s41239-019-0155-0

Jones, K. M. L., Asher, A., Goben, A., Perry, M. R., Salo, D., Briney, K. A., & Robertshaw, M. B. (2020). “We’re being tracked at all times”: Student perspectives of their privacy in relation to learning analytics in higher education. *Journal of the Association for Information Science and Technology*, *71*(9), 1044–1059. https://doi.org/10.1002/asi.24358

Jones, K. M. L., Rubel, A., & LeClere, E. (2020). A matter of trust: Higher education institutions as information fiduciaries in an age of educational data mining and learning analytics. *Journal of the Association for Information Science and Technology*, *71*(10), 1227–1241. https://doi.org/10.1002/asi.24327

Karaoglan Yilmaz, F. G., & Yilmaz, R. (2020). Student opinions about personalized recommendation and feedback based on learning analytics. *Technology, Knowledge and Learning*, *25*, 753–768. https://doi.org/10.1007/s10758-020-09460-8

Khalil, M., & Ebner, M. (2015). Learning analytics: Principles and constraints. *EdMedia 2015–World Conference on Educational Media and Technology*, 1789–1799. https://doi.org/10.13140/RG.2.1.1733.2083

Lederman, D. (2018, September 5). Unizin chooses new leaders, launches data platform. *Inside Higher Ed*. https://www.insidehighered.com/digital-learning/insights/2018/09/05/unizin-chooses-new-leaders-launches-data-platform

Liberty Justice Center. (2020, October). *Students file federal lawsuit against Indiana University over privacy violations, breach of contract*. https://ljc-assets.s3.amazonaws.com/2020/10/Gutterman-v.-Indiana-University-Press-Release_Final.pdf

Morrison, S., & Heilweil, R. (2020, December 18). How teachers are sacrificing student privacy to stop cheating. *Vox*. https://www.vox.com/recode/22175021/school-cheating-student-privacy-remote-learning

Nevaranta, M., Lempinen, K., & Kaila, E. (2020). Students’ perceptions about data safety and ethics in learning analytics. *Proceedings of the Conference on Technology Ethics*, 23–37. http://ceur-ws.org/Vol-2737/FP_2.pdf

Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.

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

Prinsloo, P. (2020). Of 'black boxes' and algorithmic decision-making in (higher) education – A commentary. *Big Data & Society*, *7*(1), 1–6. https://doi.org/10.1177/2053951720933994

Prinsloo, P., & Slade, S. (2017a). Big Data, higher education and learning analytics: Beyond justice, towards an ethics of care. In B. Kei Daniel (Ed.), *Big Data and Learning Analytics in Higher Education: Current Theory and Practice* (pp. 109–124). Springer International Publishing.

Prinsloo, P., & Slade, S. (2017b). Ethics and learning analytics: Charting the (un)charted. In C. Lang, G. Siemens, A. Wise, & D. Gašević (Eds.), *The Handbook of Learning Analytics* (pp. 49–57). Society for Learning Analytics Research (SoLAR). https://doi.org/10.18608/hla17

Prinsloo, P., & Slade, S. (2018). Student consent in learning analytics: The devil in the details? In J. Lester, C. Klein, A. Johri, & H. Rangwala (Eds.), *Learning Analytics in Higher Education: Current Innovations, Future Potential, and Practical Applications* (pp. 118–139). Routledge. http://oro.open.ac.uk/55361/

Prinsloo, P., & Slade, S. (2013). An evaluation of policy frameworks for addressing ethical considerations in learning analytics. *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, 240–244. https://doi.org/10.1145/2460296.2460344

Regan, P. M. (2000). Privacy and the common good: Implications for public policy. In *Legislating privacy: Technology, social values, and public policy* (pp. 212–243). University of North Carolina Press.

Regan, P. M., & Jesse, J. (2018). Ethical challenges of edtech, big data and personalized learning: Twenty-first century student sorting and tracking. *Ethics and Information Technology*, 1–13. https://doi.org/10.1007/s10676-018-9492-2

Reidenberg, J. R., & Schaub, F. (2018). Achieving big data privacy in education. *Theory and Research in Education*, *16*(3), 263–279. https://doi.org/10.1177/1477878518805308

Roberts, L. D., Howell, J. A., Seaman, K., & Gibson, D. C. (2016). Student attitudes toward learning analytics in higher education: "The Fitbit version of the learning world." *Frontiers in Psychology*, *7*, 1–11. https://doi.org/10.3389/fpsyg.2016.01959

Rubel, A., & Jones, K. M. L. (2016). Student privacy in learning analytics: An information ethics perspective. *The Information Society*, *32*(2), 143–159. https://doi.org/10.1080/01972243.2016.1130502

Sclater, N. (2016). Developing a code of practice for learning analytics. *Journal of Learning Analytics*, *3*(1), 16–42. https://eric.ed.gov/?id=EJ1126786

Selwyn, N. (2013). *Distrusting educational technology: Critical questions for changing times*. Routledge. https://www.routledge.com/Distrusting-Educational-Technology-Critical-Questions-for-Changing-Times/Selwyn/p/book/9780415708005

Selwyn, N. (2015). Data entry: Towards the critical study of digital data and education. *Learning, Media and Technology*, *40*(1), 64–82. https://doi.org/10.1080/17439884.2014.921628

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1510–1529. https://doi.org/10.1177/0002764213479366

Slade, S., Prinsloo, P., & Khalil, M. (2019). Learning analytics at the intersections of student trust, disclosure and benefit. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 235–244. https://doi.org/10.1145/3303772.3303796

Solove, D. J. (2008). *Understanding privacy*. Harvard University Press. http://www.hup.harvard.edu/catalog.php?isbn=9780674035072

Tene, O., & Polonetsky, J. (2013). Big data for all: Privacy and user control in the age of analytics. *Northwestern Journal of Technology and Intellectual Property*, *11*(5), 239–273. https://scholarlycommons.law.northwestern.edu/cgi/viewcontent.cgi?article=1191&context=njtip

Tsai, Y.-S., Whitelock-Wainwright, A., & Gašević, D. (2020). The privacy paradox and its implications for learning analytics. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 230–239. https://doi.org/10.1145/3375462.3375536

Tufekci, Z. (2020, September 4). The pandemic is no excuse to surveil students. *The Atlantic*. https://www.theatlantic.com/technology/archive/2020/09/pandemic-no-excuse-colleges-surveil-students/616015/

University of California-Berkeley. (n.d.). *Learning data principles*. https://rtl.berkeley.edu/learning-data-principles

University of Hawai'i at Mānoa. (n.d.). *Resolution supporting learner data privacy principles and practices*. https://drive.google.com/file/d/1TE7M7CIjibJWPqsGdBD6mBRf59D6Ldqv/view

Vogel, V. (2015, May 11). The chief privacy officer in higher education. *EdSurge*. https://er.educause.edu/articles/2015/5/the-chief-privacy-officer-in-higher-education

Vu, P., Adkins, M., & Henderson, S. (2020). Aware, but don't really care: Students' perspective on privacy and data collection in online courses. *Journal of Open Flexible and Distance Learning*, *23*(2), 42–51. http://jofdl.nz/index.php/JOFDL/article/view/350

West, D., Luzeckyj, A., Searle, B., Toohey, D., Vanderlelie, J., & Bell, K. R. (2020). Perspectives from the stakeholder: Students' views regarding learning analytics and data collection. *Australasian Journal of Educational Technology*, *36*(6), 72–88. https://doi.org/10.14742/ajet.5957

Zimmerman, J. (2020, March 10). Coronavirus and the great online-learning experiment. *The Chronicle of Higher Education*. https://www.chronicle.com/article/coronavirus-and-the-great-online-learning-experiment/

Learning Analytics without personal data? It's possible!

Thomas Dondorf
RWTH Aachen University
dondorf@lfi.rwth-aachen.de

Malte Persike
RWTH Aachen University
persike@cls.rwth-aachen.de

Heribert Nacken
RWTH Aachen University
nacken@lfi.rwth-aachen.de

ABSTRACT: The adoption of Learning Analytics in the higher education sector has stagnated in recent years. Research suggests that data privacy issues play a central role in the use of Learning Analytics. In this paper, we present RWTHanalytics, an open-source, privacy-focused software that logs no personal data while providing a user interface for descriptive course analytics. During the summer semester 2020, 63 teachers used the software in over 150 courses at two universities. One of the biggest German universities participated and more than 50,000 students generated over 50 million logged data entries. The following evaluation with 43 teachers shows great acceptance and indicates the need for basic descriptive analytics.

Keywords: Learning Analytics, Responsible Learning Analytics, Moodle, Data Privacy

1 INTRODUCTION

The adoption of Learning Analytics in the higher education sector is scarce (Williamson, 2017). While recent research trends in Learning Analytics focus on machine learning and recommendation systems (Bodily & Verbert, 2017), the higher education sector struggles to deliver basic descriptive course statistics to teachers and students due to missing privacy-friendly solutions. A recent interview study by Ifenthaler et al. (Dirk Ifenthaler & Jane Yau, 2019) comes to the conclusion that further resources are required to adopt Learning Analytics. One key issue is the handling of personal data and concerns of stakeholders regarding privacy and ethics (Drachsler & Greller, 2016).

To help increase the adoption of Responsible Learning Analytics, we created a privacy-friendly open source software solution that is easy to integrate into existing Moodle courses. The stored data cannot be traced back to individual users and due to the source code being published as open source all used algorithms can be checked and verified by users. The presented visualizations show metrics like resource usage and course accesses over time to teachers and students.

In the remainder of this paper, we first present related literature and tools. After that, we present the implementation including the system architecture and user interface of RWTHanalytics. Following, we outline our evaluation. Finally, we conclude with an outlook on the future of RWTHanalytics as well its adoption.

2 BACKGROUND

Moodle is the one of the most used learning management systems in the higher education sector. Even compared to commercial solutions, Moodle has built itself a reputation among learning management systems. But although Moodle comes with plenty of features supporting analytics in general, there currently exists no single solution that can serve as a first step to provide privacy-friendly analytics to teachers and students. Therefore, we developed RWTHanalytics[1], a software that integrates into Moodle and focuses on data privacy and practical adoption in educational institutions.

An extensive analysis of existing Learning Analytics tools for Moodle was performed before the start of development. Our literature review found multiple tools that have not received updates for several years and are not working anymore as well as literature that does not provide the used software. Other solutions we discovered included offline analysis tools and software without a user interface for teachers and students not suited for our requirements.

Moodle itself comes with built-in support for descriptive analytics named *Course reports* as well as a machine-learning system for predictive analytics (Olivé et al., 2018) called *Moodle Analytics*. An interesting Analytics plugin for Moodle we found was *Analytics Graphs*[2]. Unfortunately, the built-in solutions as well as the plugin use the built-in logging infrastructure in Moodle (Furukawa et al., 2017) which logs personalized data and cannot be considered data privacy-friendly. In summary, no software satisfied the need for privacy-respecting Learning Analytics in Moodle.

3 IMPLEMENTATION

Our software implementation was built on a value-sensitive design process. The data privacy and transparency requirements were based on current research results (Drachsler & Greller, 2016; Griffiths et al., 2016; Hoel & Chen; Steiner et al., 2016). The tool was developed in cooperation with the data privacy officer of the university. He informed the development team about legal requirements and helped to create the data privacy statement. In addition to the legal requirements, multiple requirements were added to maximize transparency towards students. The development has been iterated multiple times prior to the start of the pilot phase to incorporate feedback from all stakeholders (Dondorf et al., 2019).

The resulting software respects all legal data privacy regulations (German and European privacy law). The software logs what kind of action is happening (e.g. “Download of PDF”) at what time, but does not store who triggered the action. Therefore, the stored data cannot be traced back to individual users. The tool integrates into the Moodle user interface and uses data from the internal Moodle database for visualizations. All data is shown in a privacy-respecting way. Aggregated datasets are only shown if the set contains at least ten data points.

[1] https://github.com/rwthanalytics/moodle-local_learning_analytics - All source code is published as open-source under the GNU General Public License online

[2] https://github.com/marceloschmitt/moodle-block_analytics_graphs

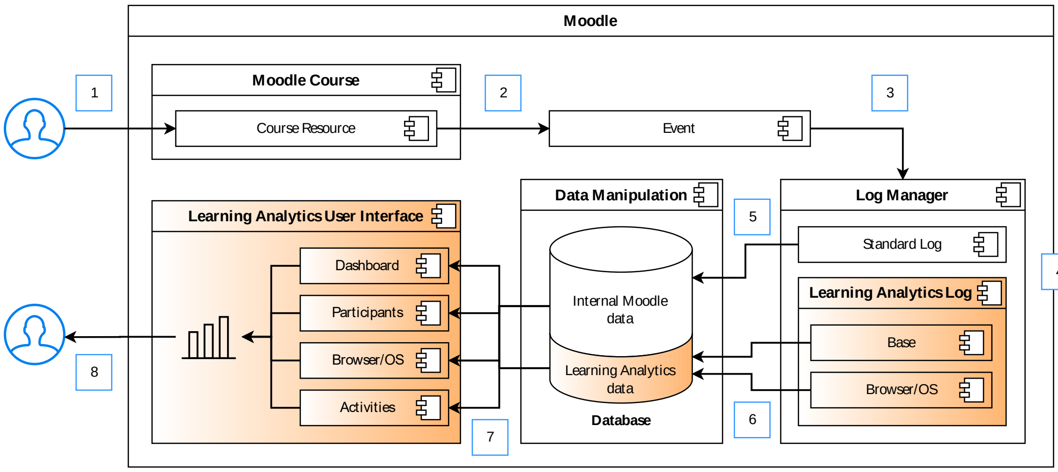
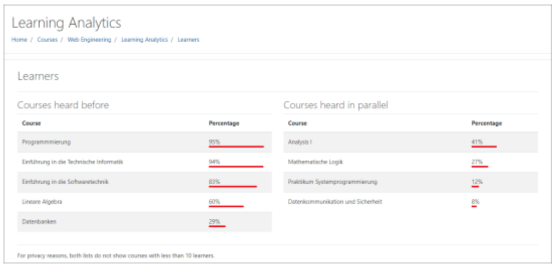


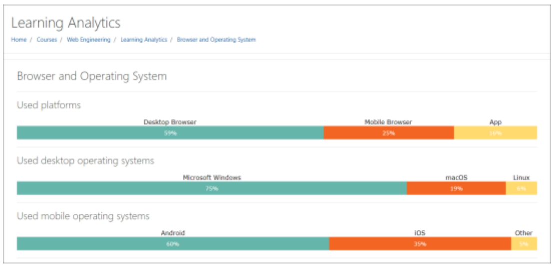
Figure 1 Architecture of the software and integration into Moodle

The overall architecture is depicted in Figure 1. Components and data of our implementation are visualized with an orange gradient. Moodle core functionality is depicted as white. The arrows show the flow of data from user action (1) to visualization (8).

(a) Dashboard



(b) Participants



(c) Browser and Operating System

(d) Activities

Figure 2 Screen captures of the pages (excluding page navigation, header and footer)

The user interface was based on a Learning Analytics dashboard by Kuzilek et al. (Kuzilek et al., 2015). The entry page features a dashboard that contains a plot at the top showing the aggregated access numbers by week. Below that, a preview for each linked page is shown containing a summary of the information from each page. The tool is integrated into the course navigation making it easy to access. Visualizations of the pages are shown in Figure 2. Figure 2a shows the dashboard. The "Participants" page in Figure 2b lists how many students were in rolled in which courses before the current course started, as well as how many students of the current course are enrolled in other courses in the same semester. Technical information about used browsers and operating systems is shown in Figure 2c. Figure 2d shows the page "Activities" that lists how often each course activity was accessed.

4 EVALUATION

During the summer semester 2020, the software was used at two German universities, the *RWTH Aachen University* (RWTH) and the *Hochschule Ruhr West – University of Applied Sciences* (HRW). The semester started in April 2020 and ended in September. Due to Covid-19, both universities changed their lectures from regular face-to-face to video-based and blended learning settings. The evaluation consisted of two parts: A system evaluation that includes technical feedback from operation and a user evaluation focused on teachers, conducted via questionnaire at the end of the semester.

The system evaluation concluded that the software was able to handle tens of thousands of users while keeping the storage size to a minimum. In total, over 44 million log entries were generated at RWTH resulting in a database size of 2.6 GB. In the following, we focus on the user evaluation.

Table 1 lists the number of participating teachers and courses. In addition, we list the number of enrollments in the participating course as well as the number of total students. Both universities operated the software in a setting that logged all events into the database, but the user interface was unlocked only for the participating courses. In total, 63 teachers with 169 courses participated. At the end of the semester, a questionnaire was sent to all participating teachers of the pilot phase. 43 of 63 teachers (68%) completed it.

Table 1 Information about participating universities in the pilot phase

University	Participating teachers	Participating courses	Enrollments in courses[3]	Total students at university
RWTH	*31*	*59*	*13,106*	*45,628*
HRW	*32*	*110*	*10,516*	*~6,500*
Total	*63*	*169*	*23,622*	*~52,000*

As teachers had to contact the service teams of each university to participate, we asked teachers how they became aware of the Learning Analytics service. We exclude the results from HRW, as the software was centrally announced via mail to all teachers there. The results show that 48% of teachers at RWTH actively asked for an analytics service. 29% were informed about the software by another teacher. 10% were informed through an internal Moodle training course of the university. The remaining 13% of teachers mentioned other reasons. The high number of teachers actively asking for Learning Analytics can be explained by Covid-19 and shows the desperate need for simply descriptive analytics. Teachers reported back to us via free text that they were desperately looking for any kind of analytics as it was hard for them to estimate how many students were following their lectures.

To assess the usability of our software, we asked the teachers to complete the 10-item System Usability Scale (SUS) by Brooke (Brooke, 1996). It received an overall average SUS score of 80.2. Using the acceptability ranges described by Bangor et al. (2009), our software ranks in the best possible range ("Acceptable"), the equivalent of a "B" on a standard school grading scale (starting at 80).

[3] The total number of enrollments for participating courses was counted, not the unique number of students. For that reason, this number can be higher than the total of students.

A question how often Learning Analytics was accessed during the semester received mixed feedback. 34% said that they used the tool once per week or even more often. 29% answered they used it once per month. 37% said they used it a few times during the semesters. The answers “in total once” and “not at all” were not selected once.

Table 2 Average rating of usefulness and standard deviation for each page

Page	Rating of usefulness	Std. dev.
Dashboard (Figure 2a)	*4.5*	*0.8*
Participants (Figure 2b)	*3.2*	*1.2*
Browser / Operating System (Figure 2c)	*2.2*	*1.2*
Activities (Figure 2d)	*4.5*	*0.7*

Regarding the user interface, each page was graded for usefulness on a 5-point scale from “Not useful” (1) to “Very useful” (5). Table 2 shows the results. The dashboard and activities functionality were received as rather useful while technical information about browser and operating system was deemed as not useful.

Further questions on a standard 5-point Likert (1-5) asked if our tool supported them in their teaching (3.6), whether teachers were interested in personalized analytics (4.8) and if they would like to compare students directly (3.6). The last question asked if they were overall satisfied with the software (3.9). An additional free text field at the end asked for open feedback. Of 20 responses, 8 teachers asked for personalized analytics, 7 expressed their gratitude for providing the software and 3 asked for statistics regarding video lectures.

5 CONCLUSION

This paper presented RWTHanalytics, an open-source, privacy-focused data analytics tool for Moodle. Based on previous research we developed a software solution built on a value-sensitive design process incorporating data privacy and transparency as central requirement. To increase adoption in other educational institutions the system was designed as plugin for the Moodle learning management system. The created user interface provides live statistics based on an extendible plugin structure.

During the summer semester 2020, 63 teachers in 169 courses used the software during a pilot phase. Overall, more than 50,000 students triggered over 50 million database log entries. The solution has proven to be scalable even in high load scenarios with ten thousands of students. The user evaluation led to mostly positive feedback and revealed interest in personalized analytics.

The results from the evaluation have led to another iteration of the software. The public release in the Moodle plugin store is currently being prepared. During the writing of this paper, a third German university has installed the software. Four more universities have expressed interest and are currently considering the use of the tool.

By providing a data privacy-friendly implementation that is easy to install and administrate, we hope to increase the long-term adoption of Learning Analytics. RWTHanalytics shows that it is possible to design a Learning Analytics software in a privacy-friendly way while still providing a real benefit to its users. We do not envision our tool as a solution for advanced Learning Analytics, but rather as a first entry point for educational institutions struggling to offer privacy-friendly analytics.

REFERENCES

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of Usability Studies*, *4*(3), 114–123.

Bodily, R., & Verbert, K. (2017). Trends and Issues in Student-Facing Learning Analytics Reporting Systems Research. In *LAK '17, Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 309–318). Association for Computing Machinery. https://doi.org/10.1145/3027385.3027403

Brooke, J. (1996). SUS: a "quick and dirty" usability scale. *Usability Evaluation in Industry*, 189--194.

Dirk Ifenthaler, & Jane Yau (2019). Higher Education Experts' Views on Learning Analytics Policy Recommendations for Supporting Study Success: A German Case. In Johan Van Braak, Mark Brown, Lorenzo Cantoni, Manuel Castro, Rhonda Christensen, Gayle V. Davidson -Shivers, Koen DePryck, Martin Ebner,Mikhail Fominykh, Catherine Fulford, Stylianos Hatzipanagos, Gerald Knezek,Karel Kreijns, Gary Marks, Erkko Sointu, Elsebeth Korsgaard Sorensen, Jarmo Viteli, Joke Voogt, Peter Weber, Edgar Weippl, Olaf Zawacki-Richter Theo Bastiaens (Ed.), *Proceedings of EdMedia + Innovate Learning 2019* (pp. 745–753). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/p/210073

Dondorf, T., Pyka, C., Gramlich, R., Sewilam, H., & Nacken, H. LEARNING ANALYTICS SOFTWARE IMPLEMENTATION FOR THE MOODLE LEARNING MANAGEMENT SYSTEM. In *Proceedings of the 12th annual International Conference* (pp. 6957–6964). https://doi.org/10.21125/iceri.2019.1655 (Original work published 2019)

Drachsler, H., & Greller, W. (2016). Privacy and Analytics: It's a DELICATE Issue a Checklist for Trusted Learning Analytics. In *LAK '16, Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (pp. 89–98). Association for Computing Machinery. https://doi.org/10.1145/2883851.2883893

Furukawa, M., Yamaji, K., Yaginuma, Y., & Yamada, T. (2017). Development of learning analytics platform for OUJ online courses. In *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*.

Griffiths, D., Drachsler, H., Kickmeier-Rust, M., Steiner, C., Hoel, T., & Greller, W. (2016). Is privacy a show-stopper for learning analytics? A review of current issues and solutions. *Learning Analytics Review*(6).

Hoel, T., & Chen, W. Implications of the European data protection regulations for learning analytics design. In *Workshop paper accepted for presentation at The International Workshop on Learning Analytics and Educational.*

Kuzilek, J., Hlosta, M., Herrmannova, D., Zdrahal, Z., Vaclavek, J., & Wolff, A. (2015). OU Analyse: analysing at-risk students at The Open University. *Learning Analytics Review*, *LAK15-1*, 1–16. http://oro.open.ac.uk/42529/

Olivé, D. M., Huynh, D. Q., Reynolds, M., Dougiamas, M., & Wiese, D. (2018). A supervised learning framework for learning management systems. In *DATA '18, Proceedings of the First International Conference on Data Science, E-Learning and Information Systems* (pp. 1–8). Association for Computing Machinery. https://doi.org/10.1145/3279996.3280014

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. https://doi.org/10.18608/jla.2016.31.5

Williamson, B. (2017). *Big data in education. The digital future of learning, policy and practice*. Sage.

Calling for a More Responsible Use of Student Data in K-12 Education

Olga Viberg
KTH Royal Institute of Technology, Sweden
oviberg@kth.se

Annika Andersson
Örebro University, Sweden
Annika.andersson@oru.se

Ella Kolkowska
Örebro University, Sweden
Ella.kolkowska@oru.se

Stefan Hrastinski
KTH Royal Institute of Technology, Sweden
stefanhr@kth.se

ABSTRACT: This positioning paper raises several critical issues related to the protection of children's privacy in K-12 education and calls for a need for more empirical research in this regard. In this paper, we argue that the protection of children's privacy in K-12 schools is a moral and legal obligation for school leadership, teachers and parents. In light of the ongoing pandemic, this topic is especially critical with schools worldwide having to integrate digital technologies to be able to teach. The use of technologies comes with increased access to student data where students' actions and behaviours are continuously traced. The collection and analysis of such data is often crucial for improving teaching and learning but it can also be misused. Mixing various digital products, re-analysing and merging different data sets increase the risk of revealing sensitive student data. Protection of privacy is legally regulated and violation of it has serious legal consequences for schools. Protection of children's privacy online in schools is also a moral obligation, since children lack agency in protecting their personal data with parents, teachers and school leaders being the main guardians. Schools must not create new privacy risks for children by acting irresponsibly in online learning settings.

Keywords: Student data, responsible use, K-12 education, privacy, children, learning Analytics

1 INTRODUCTION

Schools worldwide have integrated digital technology into K-12 education (Macarini et al., 2019; Vezzoli, Mavrikis, & Vasalou, 2020). Such use offers increased access to personal and interaction student data but also creates new risks for students' privacy (Lieberman, 2020), including concerns about how schools monitor students' technology use, what data companies get when students use their devices, and how schools use the collected data (Kumar et al., 2019). New concerns regarding student privacy (Bulger, McCormick, & Pitcan, 2017; Kumar et al., 2019; Livingstone, 2020) are also emerging in relation to learning analytics (LA) that refers to the collection, measurement, analysis and use of student data for improving teaching and learning, and the contexts in which they occur (Long & Siemens, 2011).

The new trends and the move to remote teaching (also recently imposed by the pandemic) opened up for increased access not only to more data, but also learning contexts that were not earlier in the 'data gaze' (Beer, 2019), which refers to how data-driven and data-led analytics continue to expand in its reach, but also how the 'gaze' intensifies in its ability to reveal what has, up to now, been hidden – of learning management system providers, intra-institutional learning platforms, and commercial interests (Williamson, Eyon, & Potter, 2020). This suggests that we now have access to not only more data, but also to more nuanced data from many sources, often in real-time, allowing schools and companies detailed student profiles creating so-called 'a datafied child' (Lupton & Williamson, 2017; Williamson et al., 2020). 'A datafied child' refers to how intersecting monitoring platforms and devices use children's data to create digital profiles that may accompany children long after the period of collection. Children's records, consisting of demographic, behavioral and relational data, are stored, combined with other information sources and accompany them over the years. Such data is increasingly shared and sold to commercial interests as part of service level agreements and practices that breach students' right to privacy with possible long-term impacts (Regan & Jesse, 2019). In sum, there are different parties that are interested in using student data for various purposes, but the question is *who will protect the 'datafied' child's right to privacy in this context?* To protect children's privacy in schools is a moral and legal obligation for school leadership, teachers and parents. Yet, there is very little research about responsible use of students' data in schools (Kumar et al., 2019).

2 CHILDREN'S RIGHTS TO PRIVACY

As stressed by Livingstone (2020), "the status of the child online is shifting from one of invisibility to one of hypervisibility in an increasingly datafied world, and the child's right to privacy has rapidly become the most contested of all the rights" (p.1). Consequently, in the context of the many challenges encountered by K-12 educational institutions, and evidence informed decision-making, the promise and concerns regarding data analytics, student agency and student privacy are pertinent. The increased datafication of K-12 education is understood in the broader context where offline and online behaviors are converted "into online quantified data, thus allowing for real-time tracking and predictive analysis" (Van Dijck, 2014, p. 198). Various forms of dataveillance (Albrecthslund & Lauritsen, 2013; Lupton & Williamson, 2017; Van Dijck, 2014) entail the collection of data, increasingly real-time, from children which are stored in online corporate platforms, calculated so as to predict and manipulate future behaviour, and therefore monetised (Hintz et al., 2017; Mascheroni, 2018; Van Dijck, 2014, cited in Mascheroni, 2018, p.3). Lupton and Williamson (2017) suggest that hitherto there is little evidence of specific consideration to safeguard children's rights in relation to dataveillance and propose paying attention to this. Furthermore, Kumar et al. (2019) highlight that "little research has examined how educators consider privacy and security in relation to classroom technology use" (p. 1). There is also, to our knowledge, a paucity of empirical research that investigates the school leadership- and parents' perspective in this regard. All these perspectives are important to be able to *define* and *enact* a more responsible use of student data in K-12 education settings (i.e., in practice).

A responsible use of children's data means the protection of their data privacy rights, which has received special attention in several legislative documents, such as the General Data Protection Regulation (GDPR) in the European context and the Family Educational Rights and Privacy Act (FERPA) in the context of the US, and also specific toolkits, such as General Principles of Children's Online Privacy and Freedom of Expression (UNESCO, 2018). However, these documents do not accurately reflect different stakeholders' perspectives and do not provide practical guidelines about how they should be implemented in practice - in particular educational contexts - in which different cultural values may also play a role. For example, in the context of the Nordic countries, such values as trust, openness and transparency are ingrained in the culture at different levels, including the individuals' attitudes towards- and their acceptance of the use of emerging digital technologies (Robinson, 2020), for example artificial intelligence (AI) that is nowadays increasingly used in the design of adaptive and personalized learning solutions that have being extensively adopted and used in education in different forms (Chen et al., 2020). Yet, most of the AI-based educational tools have not been built with the protection of students' privacy in mind (Rauf, 2020).

In her work, Livingstone (2020) poses several incitements or provocations to be considered in terms of protecting children's privacy. They include: 1. Should we enable children's full participation in digital spaces, or should we minimize their risks by providing children-only, or even off-line, spaces? 2. Should we pay from the public funds for the provision of digital and non-digital spaces for children, or should we allow the commercialisation of these spaces? 3. Is it an option to hold parents accountable? In this regard, she points to evidence that parents are frequently not up to the task, and that the parents of the most vulnerable are equally not prepared to take accountability. In their efforts to protect children in online spaces, they may interfere with children's rights of participation in digital spaces. 4. To what extent can we hold digital companies responsible for the digital well-being of children? And even if we do, are they trustworthy? These are important questions to reflect upon in the design of relevant research studies in various cultural and socio-economical contexts, as well as in the process of examining privacy aspects in the evaluation of the digital technologies use in K-12 education settings.

3 PROBLEMS TO ADDRESS

Overall, future research aiming to define and enable a more responsible use of student data in K-12 education should address the following problems:

The *first* problem relates to the fact that (educational) technology is seen as the great equaliser for different student populations and geopolitical contexts (Prinsloo, 2018). Governments, educational departments and educators worldwide are encouraged and also nowadays 'forced' to use digital technologies, often without carefully considering the appropriateness and effect of these technologies in context, i.e., consequences on student privacy in K-12 school context.

The *second* problem is that the collected data is largely seen as neutral, pre-analytic and representing an objective state of affairs. Student data is collected, often without sufficient consideration for issues surrounding privacy, security, confidentiality and downstream use of the data. Students' records consisting of demographic, behavioural, learning data, and relational data are stored, often combined with other information sources and may accompany children across their enrolment, often inter-school. These digital dossiers, their scope, and the understandings that inform the data and interventions based on these dossiers, become data-doubles that impact on students' lives, far beyond the context where the data were initially collected (Williamson, 2019). In K-12 school settings, where young students have limited data agency, these dossiers have critical implications for their options later in their lives.

The *third* problem refers to the reality that increasingly, these student data sets are shared, and even sold to commercial and third-party interests as part of service level agreements, data breaches and other practices that violate students' right to privacy, with possible long-term impacts (Regan & Jesse, 2019). If researchers and K-12 institutions continue to develop data analytics projects and infrastructures in order to improve teaching and learning the obligation to do so responsibly will increase as well. In achieving this, we need to carefully consider different stakeholders' perspectives and practices, as well as to involve them into co-design of practical ways to be implemented in schools to protect children's privacy in increasingly growing online educational settings. Milkaite and Lievens (2020), for example, suggest that possible practical ways of enhancing transparency for children (in regard to how their data is being used) should include "legal visualisation, co-design, co-creation techniques and participatory design methods which focus on presenting legal information in a transparent and clear manner" (p.5). However, such related practical efforts have hitherto been very scarce.

4 CONCLUSION

We began this paper by pointing out that there is a scarcity of research specifically considering children's rights in relation to dataveillance in K-12 education and that there is even less empirical research investigating the school leadership-, teachers- and parents' perspective in this regard. We continued by outlining several critical issues to be considered in regard to the protection of children's privacy in K-12 education settings. We concluded by calling for more empirical research that a) carefully considers the appropriateness and effect of digital technologies in context, i.e., the consequences on student privacy in K-12 school context; b) considers where the student dossiers are being used beyond the context where the data was initially collected; and c) considers different stakeholders' perspectives and practices, as well as involves them into the co-design of practical ways to be implemented in schools to protect children's privacy. If these issues were to be researched, we believe that we will be able to define and enable a more responsible use of young students' data in K-12 education.

REFERENCES

Albrechtslund, A., & Lauritsen, P. (2013). Spaces of everyday surveillance: unfolding an analytical concept of participation. *Geoforum, 49,* 310–316.

Beer, D. (2019). *The Data Gaze: Capitalism, Power and Perception (Society and Space).* Sage.

Bulger, M., McCormick, P., & Pitcan, M. (2017). The Legacy of InBloom. Working paper. Data & Society. https://datasociety.net/wp-content/uploads/2017/02/InBloom_feb_2017.pdf 2020-09-15

Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelligence in education: A review. *IEEE Access, 8*, 75264-75278. https://doi.org/10.1109/ACCESS.2020.2988510.

Hintz, A., Dencik, L., & Wahl-Jorgensen, K. (2017). Digital citizenship and surveillance society: Introduction. *International Journal of Communication 11,* 731–739.

Kumar, P.C., Chetty, M., Clegg, T., & Vitak, J. (2019). Privacy and security considerations for digital technology use in elementary schools. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3290605.3300537

Lieberman, M. (2020). Massive shift to remote learning prompts big data privacy concerns. *EducationWeek.* Accessed from https://www.edweek.org/technology/massive-shift-to-remote-learning-prompts-big-data-privacy-concerns/2020/03 2021-01-22

Livingstone, S. (2020). Realizing children's rights in relation to the digital environment. European Review. Retrieved from http://eprints.lse.ac.uk/103563/ 2021-01-22

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review, 46*(5) (2011), 31-40.

Lupton, D., & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. *New Media & Society*. https://doi.org/10.1177/1461444816686328

Macarini, L. A., Cechinel, C., Santos, H. L. D., Ochoa, X., Rodés, V., Alonso, G. E., ... & Díaz, P. (2019, March). Challenges on implementing learning analytics over countrywide K-12 data. In Proceedings of the *9th International Conference on Learning Analytics & Knowledge* (pp. 441-445). https://doi.org/10.1145/3303772.3303819

Milkaite, I., & Lievens, E. (2020). Child-friendly transparency of data processing in the EU: from legal requirements to platform policies. *Journal of Children and Media, 14*(1), 5-21, DOI: 10.1080/17482798.2019.1701055

Mascheroni, G. (2018). Datafied childhoods: Contextualising datafication in everyday life. *Current Sociology, 1-16*. https://doi.org/10.1177/0011392118807534

Prinsloo, P. (2018). Include us all! Directions for the adoption of learning analytics in the Global South: An African perspective. Digital Learning for Development (DL4D). Available at: http://dl4d.org/portfolio-items/learning-analytics-for-the-global-south/

Rauf, D. (2020). Artificial intelligence in K-12 education: Unintended consequences lurk, report warns. Retrieved from https://www.edweek.org/technology/artificial-intelligence-in-k-12-education-unintended-consequences-lurk-report-warns/2020/05

Regan, P., & Jesse, J. (2019). Ethical challenges of edtech, big data and personalized learning: Twenty-first century student sorting and tracking. *Ethics and Information Technology 21*(3), 167–179.

Robinson, S. (2020). Trust, transparency, and openness: How inclusion of cultural values shapes Nordic national public policy strategies for artificial intelligence (AI). *Technology in Society, 63.* https://doi.org/10.1016/j.techsoc.2020.101421

UNESCO (2018). General principles on children's online privacy and freedom of expression; Industry toolkit.https://sites.unicef.org/csr/files/UNICEF_Childrens_Online_Privacy_and_Freedom_of_Expression(1).pdf

Van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society 12* (2), 197-208.

Vezzoli, Y., Mavrikis, M., & Vasalou, A. (2020, March). Inspiration cards workshops with primary teachers in the early co-design stages of learning analytics. In *Proceedings of the Tenth*

International Conference on Learning Analytics & Knowledge (pp. 73-82). https://doi.org/10.1145/3375462.3375537

Williamson, B. (2019). Policy networks, performance metrics and platform markets: Charting the expanding data infrastructure of higher education. *British Journal of Educational Technology.* https://doi.org/10.1111/bjet.12849

Williamson, B., Eynon, R., & Potter, J. (2020). Pandemic politics, pedagogies and practices: digital technologies and distance education during the coronavirus emergency. *Learning, Media & Technology, 45*(2). https://doi.org/10.1080/17439884.2020.1761641

Unexpected Consequences with using Transparency as guiding principle in Project-based work in Higher Education

Lena-Maria Öberg
Mid Sweden University
lena-maria.oberg@miun.se

Jörgen Söderback
Mid Sweden University
jorgen.soderback@miun.se

Thomas Persson Slumpi
Mid Sweden University
thomas.persson@miun.se

ABSTRACT: Project-based learning has been shown to have several positive effects. The course 'Distributed Software Development' was designed aiming to teach students how to develop software in a dispersed project. An initial idea was that focus should be on the working process instead of the final artefact and therefore there was a need to monitor as much of the process as possible. In order to evaluate and grade individual students' contributions to the project, teachers needed transparency. We here elaborate on some unexpected consequences when implementing transparency. We discuss three different unexpected consequences in our paper. The first one is related to that transparency could be seen as surveillance, the second is connected to both negative and positive aspects of that the students have the same level of transparency as the teachers. The last and third one is that the number of tools used to reach a high level of transparency makes it hard to get an overview of the students' performance. Based on these findings we argue that there is a need for further studies to be able to develop guidelines for both teachers and students.

Keywords: learning analytics, data, project work, transparency, ethics

1 INTRODUCTION

'Informatics with Focus on Systems Development' is a bachelor's programStudents can apply to this program either as a campus program or as a fully online program, but both student groups are studying together. The format is blended synchronous learning (Hrastinski, 2019) which means that lectures and other scheduled learning activities are held on campus and are at the same time live streamed so that students can participate in various study activities from anywhere. The program aims to educate software developers where knowledge and skills in for example computer programming, databases, requirement capture, troubleshooting etc. are important. It is also important that system developers also have 21st century skills, i.e., critical thinking, communication, collaboration, and creativity. Project-based learning has been shown to have several positive effects (Bell, 2010) and harmonizes with the professional life of a software developer where most work is done in projects. Many of the courses included in the program therefore have elements of group work and some courses are conducted entirely in project form. But there are also some negative aspects with group work that need to be addressed, for example challenges to assess and evaluate the efforts of individual students, freeloaders, free riders, and internal conflicts (Felps et al., 2006). Students can also have negative attitudes towards group work, based on earlier experiences, which can decrease their motivation etc. In the worst case, it can lead to students avoiding courses with elements of group work (Feichtner & Davis, 1984).

'Distributed Software Development' is one of the project courses mentioned above, aiming to teach students how to develop software in a dispersed agile project. The course is supposed to give students an understanding of social and technological challenges and how to deal with those. The course spans over five weeks and during the first week students are being introduced to agile project management and various tools for communication, collaboration, and co-creation online. An initial idea was that focus should be on the working process instead of the final artefact and therefore there was a need to monitor as much of the process as possible. In order to evaluate and grade individual students' contributions to the project, teachers needed transparency which was enabled by the digital tools being used by the students and the use of Scrum. Scrum is a framework for project management where transparency is a guiding principle. Transparency in this sense can also benefit learning in a way that it enables the possibility to support students when they get stuck without much delay. From the first week and on, students are working with their projects. The teachers allocate students into teams with five to seven members in each team, depending on the number of students in the class, and a teacher is assigned to each team to act as a facilitator. All teams have to use Scrum for project management. From a teacher perspective, there are many positive effects of transparency but the aim of this paper is to elaborate also on some unexpected consequences when implementing a high level of transparency. Our article is based upon earlier research and the context and the data collection is described in details in Söderback et al (2016). In this position paper, we will use quotes from those interviews to support our arguments.

1.1 Scrum

Scrum is a framework for agile project management and consists of a few rules, principles, and values which makes it easy to learn, but it has been shown that it is hard to master (Hassasni-Alaoui et al.,2020). The process is iterative and incremental and every iteration starts with a meeting where work is prioritized and divided into a manageable number of tasks that the team can complete during the current iteration (Schwaber & Beedle, 2002). Scrum implements empirical process control which means that progress is based on observations and transparency becomes an important element. Transparency is achieved by a mix of digital tools, intensive interactions, and extensive communication (Schwaber & Beedle, 2002). A Scrum team consists of three to nine developers and one Scrum master. A Scrum master is more like a facilitator than a traditional project leader, and the main responsibilities are to remove all obstacles for the team and make sure that the project adheres to the Scrum framework (Schwaber & Beedle, 2002). Students are developers and teachers have the roles of Scrum masters in this course. A team usually starts every day with a short meeting where every developer (student) answer the three following questions: "What did you do yesterday? What will you to today? "What is blocking your progress?" These daily meetings are important in that they create transparency and informs the team about the progress of the project and every team member is obliged to attend.

Scrum teams are self-organized and cross-functional (Schwaber & Beedle, 2002) so students are responsible for planning and organizing their work, as long as they don't violate the principles and values in the Scrum framework. Teams need to use tools for communication, project management, and software development, code sharing, versioning, communication, online meetings, planning, organizing, co-writing and co-creation, and keeping track of the progress of the project. A few of the tools are mandatory, chosen by the teachers to support different aspects of the projects and at the same time create transparency and openness, which are important elements in Scrum. Every team member is supposed to pick one task at a time and work on that task until it is completed. The Scrum board is a simple but a clear way to visualize the progress. The board shows what tasks that has to be done, tasks that are in progress, and tasks that are completed, illustrated by cards divided into three columns; from left 'to do', 'doing, and 'done'. Every card contains information needed to complete the task. Team members can follow the progression of the project and the development of their artefact. Every team works together with a fictitious customer who informs team members about system requirements.

1.2 Data sources and the ground for transparency

During a Scrum project, a lot of data is created, and since this was used in a learning situation some of the data sources are extra just to be able to follow each student individually. The transparency makes it possible to monitor the progress of the project, but it also enables surveilling of individual team members. Following data are created by team members and are possible to monitor by everyone in the team, teachers included. The amount of source code that are written by each student over time, which and how many tasks a student has accomplished over time, what every student is working on for the moment, attendance to meetings, written communication, team meeting notes. As a complement to these data, each student is also writing a diary, an individual evaluation, and an individual team evaluation document.

2 LEARNING ANALYTICS, TRANSPARENCY, AND SURVEILLANCE – FROM A THEORETICAL PERSPECTIVE

Transparency has been identified as a crucial principle for learning analytics adoption and is also the principle with the greatest number of concerns (Dringus, 2012; Lawson et al., 2016; Pardo and Siemens, 2014; Slade & Prinsloo, 2013; Tsai et al., 2020). The main argument for transparency in learning analytics is that participation in digital environments is not a blanket permission for data use (Slade & Prinsloo, 2013). Students that are the main contributor of data should be informed of all the ongoing learning analytics activities including for example the type of data collected, data collection and -processing methods, data storage, data transfer, handling of historical data, student access and feedback, etc. (Pardo & Siemens, 2014; Slade & Prinsloo, 2013; Slade et al., 2019). From a theoretical viewpoint transparency is described as important to be able to use learning analytics in an ethical way. But transparency will also make it possible to monitor students' behavior and their activities, writes that workplace monitoring has been studied extensively but that there a few studies of the impacts of surveillance practices in online learning environments.

> "It may be that students interpret strong surveillance environments as expressions of caring and concern for their learning, and therefore increased monitoring will actually increase levels of student trust. " p.23 Knox (2010).

So, transparency is crucial principle for learning analytics, but this could also bring unintended consequences when it comes to surveillance.

3 LEARNING ANALYTICS, TRANSPARENCY AND SURVEILLANCE – FROM A PRACTICAL PERSPECTIVE

Transparency was one of the guiding principles when the topical course was designed. This was not based on any theoretical framework for ethics in learning analytics rather the reason for the design principle was based on one of the principles for agile software development. As described by Söderback et al (2016) Scrum implements empirical process control, where transparency is a very important component. The students share valuable information at the daily meetings, which gives teachers a good view of individual and team progression. We have identified some unintended consequences of transparency and below we discuss positive and negative experiences of those consequences.

The first consequence concerns transparency and surveillance. Already in advance, the teachers of the course discussed whether the students would see the transparency as something negative and as surveillance, e.g., to meet a teacher daily and discuss what they did yesterday and what the plan is for the day. The tools that were used would also show if the student had checked in their code or not (or just saying they had made the code). Transparency thereby gives the teachers the possibility to give valuable and timely information about the progression of the students and thereby could give the students relevant feedback. In the evaluation that was performed after the course, the students got a chance to reflect upon this and were asked if they felt monitored (surveilled). The students had a very positive attitude to the course, and they generally describe the transparency as something positive. Knox (2010) present ideas that the relation between surveillance and trust could be different in a learning situation compared to situations at work. This could explain why the students didn't think of the transparency as surveillance and instead saw it as an expression for caring as stated by Knox (2010). The students describe the transparency, as a good thing since it gave them an overview of the other team members' tasks. One example of this is the following quote from a student:

> "I felt a little bit nervous from the beginning since the teacher would "see everything", but it quickly subsided when the course started. "

The study didn't involve all students and since Söderback et al (2016) did their study, no data have been collected that concern the student's reflections regarding transparency. During the years there are a few students that have chosen not to take this course and it not clear what kind of students that make this decision and the reasons behind it.

A second consequence is related to that all members of the team can see what is going on in the team. So – if one of the students have problems with completing their tasks, the other team members will know about it. In the agile method, this is important since it could lead to re-allocation of tasks or resources. In the study performed by Söderback et al (2016) several students described that they saw this as something positive. In earlier project, they have spent a lot of time trying to understand what to do next and what the others in the team had done. But there are several possible consequences of this. It might be a problem if some of the students do not reach the learning objective (based on their contributions being too limited). There might also be disagreements when students aiming for higher grades want to do more tasks than their team members. Finally, it could be very stressful for students that find the course hard that all the others are aware of your problems. The following quote is one example of this:

"I think that it has been good. That you do not have to think about ..what is that person doing right now. There is some kind of clearness and then you also know which one you could ask if you can see someone that is doing something similar as I am. So, I can see what Fredrik is doing. On the other hand, you cannot drop off and do something else..."

A third unexpected consequence is related to the fact that to be able to run this course the department use a number of different tools, such as boards, code-sharing tools, communication tools. This makes it hard for the teachers and the students to get at good overview and there is a risk that it is not as transparent as intended. This could lead to decisions being made on feeble basis, for example when it comes to the grading process.

To sum up we think that transparency can be a solution for learning analytics in project work in higher education, but there is a need for further studies to be able to develop clear guidelines both for teachers and students. It has been several years since Söderback et al. (2016) performed their study and based on the development within the area of personal data and for example self-tracking (see for example Fors et al., 2020) it is important to acknowledge that the student's perception of transparency might change over the years. This means that it is important to find a way to keep the students aware and updated on how and why transparency have been implemented in a course. An important step in this research is to perform studies that could give more insights in how the student perceive transparency. In what kind of courses is it suitable? Are there any group of students that dislike transparency more than other groups? How could we inform the students about how and why transparency is an important technique?

REFERENCES

Brown, M., & Klein, C. (2020). Whose data? which rights? whose power? A policy discourse analysis of student privacy policy documents. *The Journal of Higher Education*, *91*(7), 1149-1178.

Bell, S. (2010). Project-based learning for the 21st century: Skills for the future. *The clearing house*, *83*(2), 39-43.

Dringus, L. P. (2012). Learning analytics considered harmful. *Journal of Asynchronous Learning Networks*, *16*(3), 87-100.

Feichtner, S. B., & Davis, E. A. (1984). Why some groups fail: A survey of students' experiences with learning groups. *Organizational Behavior Teaching Review*, *9*(4), 58-73.

Felps, W., Mitchell, T. R., & Byington, E. (2006). How, when, and why bad apples spoil the barrel: Negative group members and dysfunctional groups. *Research in organizational behavior*, *27*, 175-222.

Fors, V., Pink, S., Berg, M., & O'Dell, T. (2020). *Imagining personal data: Experiences of self-tracking*. Routledge.

Hassani-Alaoui, S., Cameron, A. F., & Giannelia, T. (2020, January). "We use scrum, but…": Agile modifications and project success. In *Proceedings of the 53rd Hawaii International Conference on System Sciences*.

Hrastinski, S. (2019). What do we mean by blended learning?. *TechTrends*, *63*(5), 564-569.

Knox, D. (2010). A good horse runs at the shadow of the whip: Surveillance and organizational trust in online learning environments. *The Canadian Journal of Media Studies,* 7, 07-01.

Lawson, C., Beer, C., Rossi, D., Moore, T., & Fleming, J. (2016). Identification of 'at risk'students using learning analytics: the ethical dilemmas of intervention strategies in a higher education institution. *Educational Technology Research and Development*, *64*(5), 957-968.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, *45*(3), 438-450.

Schwaber, K., & Beedle, M. (2002). *Agile software development with Scrum* (Vol. 1). Upper Saddle River: Prentice Hall.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1510-1529.

Slade, S., Prinsloo, P., & Khalil, M. (2019, March). Learning analytics at the intersections of student trust, disclosure and benefit. In *Proceedings of the 9th International Conference on learning analytics & knowledge*, 235-244.

Söderback, J., Hrastinski, S., & Öberg, L. M. (2016). Using distributed scrum for supporting an online community-a qualitative descriptive study of students' perceptions. In *Proceedings of the 10th international conference on networked learning,* 397-404.

Tsai, Y. S., Perrotta, C., & Gašević, D. (2020). Empowering learners with personalised learning approaches? Agency, equity and transparency in the context of learning analytics. *Assessment & Evaluation in Higher Education*, *45*(4), 554-567.

Tsai, Y. S., Whitelock-Wainwright, A., & Gašević, D. (2020). The privacy paradox and its implications for learning analytics. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge,* 230-239.

Idiographic Learning Analytics: A Within-Person Ethical Perspective

Sonsoles López-Pernas
Universidad Politécnica de Madrid
sonsoles.lopez.pernas@upm.es

Mohammed Saqr
KTH Royal Institute of Technology, University of Eastern Finland
mmas3@kth.se

ABSTRACT: One of the main obstacles impeding the widespread use and adoption of learning analytics is the threat that it poses to students' data privacy. In this article, we present a proposal for generating person-centered insights for learners by combining data from multiple sources while preserving students' privacy. The basis of our approach is idiographic learning analytics, in which data are collected and insights are generated for each student individually. On the one hand, all the data collection and processing are performed locally on the student's device, thus preserving student privacy. On the other hand, being based on person-based methods, the idiographic approach helps deliver personalized insights.

Keywords: Ethics, Learning Analytics, Idiographic, Privacy.

1 INTRODUCTION AND BACKGROUND

Inspired by the encouraging industrial models that succeeded in converting data insights into competitive advantage, Learning Analytics (LA) were convened in 2011. LA can be defined as the "measurement, collection, analysis, and reporting of data about learners and their contexts, for the purposes of understanding and optimizing learning and the environments in which it occurs" (Siemens, 2013). Research in the field of LA started as an exploration of the potentials of using data generated by learning management systems (LMSs) to predict student performance. Nowadays, the range of applications has grown tremendously to include a wide range of methods, data sources, tools, and diverse threads of research across different fields (Saqr, 2015; Siemens, 2013). Such growth has been associated with wide-ranging interests from countries, universities, and many institutions to harness the benefits of data in education, and is expected to increase with the spike on online education with COVID-19 (Saqr & Wasson, 2020). Yet, the progress in learning analytics is closer to the research laboratories rather than real classrooms.

One of the main obstacles hampering the widespread use and adoption of LA is the threat that it poses to students' data privacy (Saqr, 2017). Indeed, the growth in LA research has not been matched by research in ethics and privacy, nor has there been enough policies developed or enacted across institutions that provide a healthy and safe ground for LA (Viberg et al., 2018). As ethics, privacy, and learners' protection lagged behind applications, so did the adoption of LA (Tsai et al., 2020). The scale of applications of LA, and the prospective future growth in such applications with ever-expanding technologies, is making it difficult for policymakers to match such pace with appropriate policies that can adapt to the vastly changing field of technology and its applications. What is more, research in LA has been more focused on institutional goals and perspectives rather than the needs and aspirations of learners themselves. Common objectives like decreasing dropout rates, and improving success rates are commonly cited as reasons for deploying LA applications (Bergdahl et al., 2020). In summary, although technology has shaped our learning and teaching, problems remain looking for a solution. LA has contributed to our understanding of learning; yet, efforts are impeded by a vastly growing field and lagging ethical policies. Therefore, a shift is needed where students are involved in the creation, understanding, and sense-making of their own data for their own sake.

As Winne has argued, the current approach of collecting large amounts of data from a group of learners (i.e., nomothetic LA) to derive insights about their behavior can hardly be generalizable (Winne, 2017). In other words, what applies to a group as an average behavior does not apply to the individual learners as each is a unique case (Molenaar & Campbell, 2009). In turn, the availability of high-resolution data generated by students enables another type of analytics where students can get just-in-time person-centered advice and support (Winne et al., 2017). This type of analytics is known as idiographic LA. This person-based approach has been gaining momentum in psychology research during the past decade. The move was kindled by increasing interest in delivering precisely personalized scientific interventions. Other fields have already benefited from idiographic approaches, e.g., precision medicine which has started to attract many researchers (Cook et al., 2018; Epskamp et al., 2018). Only recently the power of idiographic LA has been recognized as essential in uncovering the rich dynamics of cognitive development (Hofman et al., 2018).

In idiographic or person-based learning analytics, students are the data collectors, the analyzers, and the sense-makers. Data are collected from individuals with high intensity to generate enough observations, so the calculated statistics are based on many observations of a single individual, and hence the resulting mean, correlations, and predictions are of the very person (Epskamp et al., 2018; Wright et al., 2019; Saqr & López-Pernas, 2021). The abundance of data about learners from multiple sources allows such intensive data methods, e.g., data from LMSs, student information systems (SIS), and library services. This information can be complemented with data that students already have on their phones (e.g., mobility, screen time, and physical activity) and other devices of their own (e.g., fitness bands, personal computers, and tablets). The wealth of such data can be collected locally on a student's own device, analyzed locally (i.e., algorithms act solely on students' own data), and the results of such analysis can be acted upon locally as well (i.e., inferences, predictions, prescriptions, etc. can be presented exclusively to the student). In other words, the whole lifecycle of learning analytics can be performed locally, eliminating the main threats related to students' privacy.

Our proposal is a bottom-up approach to learning analytics that starts from the students, in contrast to the top-down approach commonly implemented. Such an approach has been tested and proven useful in other domains. For instance, a recent large-scale meta-analysis on consumer-based wearable activity trackers has proven that access to physical fitness dashboards on their own devices has helped individuals increase in daily step count, physical and energy expenditure (Brickwood et al., 2019). Our proposal is rooted in learning theories such as self-regulated learning that views students' agency as a fundamental element where students can influence their learning, set goals, reflect on their learning activities, and take proper action.

2 AN IDIOGRAPHIC LEARNING ANALYTICS SOLUTION TO ETHICAL CONCERNS

We propose a mobile application in which students combine data from multiple sources to generate idiographic (or person-centered) learning analytics. All the data retrieved from the allowed data sources for a single student are made available within the mobile application. Then, the data analysis is performed locally, and the results from such analysis are only presented to the student himself/herself through a dashboard in the mobile application. Figure 1 shows an overview of the solution proposed.

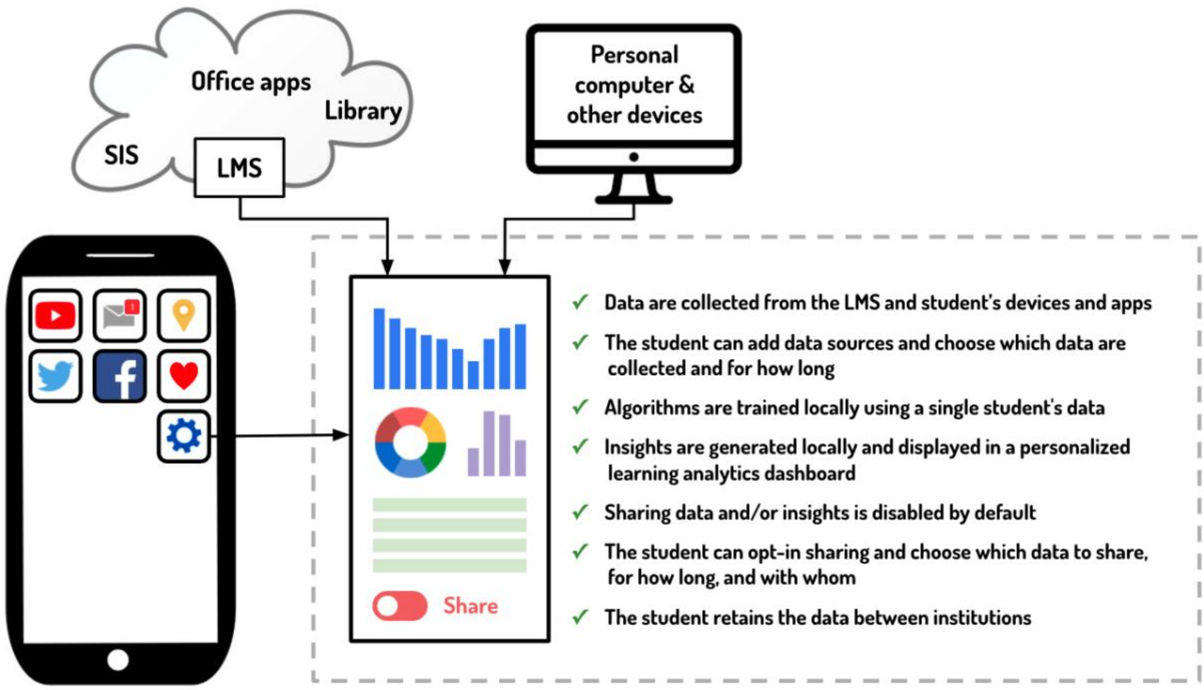


Figure 1. Proposed solution: data are collected from multiple sources and analyzed locally

2.1 Data collection

The mobile application retrieves the students' data from the LMS, including grades and performance data, as well as logs from the students' interactions with the learning materials and resources within the LMS. Students can select which supplementary data sources they wish to add as an input to enrich the insights drawn from their learning data. These auxiliary data sources can be (1) institutional services (e.g., SIS, library access, productivity applications, video conferencing apps), (2) other mobile applications and utilities (e.g., screen time, fitness, social), and (3) other devices of students' own (e.g., tablets, computers, smartwatches). All data retrieval is performed through secure encrypted connections between the mobile application and the data sources. Students can enable or disable data sources for a given time period and choose which exact data provided by a certain source they wish to make use of in their learning analytics personal dashboard.

2.2 Data analysis

The main idea behind idiographic learning analytics is that only data from a single student are operationalized. Thus, there is no need to combine data from a cohort of students to extract valuable insights about a specific learner. Without this requirement to combine data from multiple students in a centralized way, the data analysis for each student can be performed locally on each student's device. This also eliminates the need to anonymize their data since all the operations are performed locally. In our proposed solution, the complete data analysis and machine learning techniques used to, e.g., suggest learning materials or recommend a learning strategy, are applied within the mobile application using the data sources of students' choice.

2.3 Presentation and action

After data collection and analysis, the results obtained are presented to the student through a learning analytics dashboard embedded within the mobile application. In this dashboard, students can gain insights from their past performance, predicted future outcomes, and prescriptions on how to improve their academic achievement and learning strategies. The dashboard accounts for data provenance, informing the students of the exact data sources used to come up with a specific result or outcome. Although complete privacy of student data and results is enforced by default, students can share —by their own choice— specific insights with their parents, teachers, and/or peers to take their advice and guidance into consideration. Moreover, students can retain the data and generate insights as they progress throughout their educational journey, even if they change to a different institution.

3 CONCLUSIONS

In this article, we have presented a proposal for a mobile application that can generate person-centered insights for learners while preserving student privacy. The basis of our approach is idiographic learning analytics, in which insights are generated from data from a single learner (N=1). On the one hand, by means of the high-resolution data generated by students, the idiographic approach becomes essential in uncovering the rich dynamics of cognitive development (Hofman et al. 2018, Winne, 2017). On the other hand, since this approach studies each learner individually, all the data collection and processing can be performed locally on the student's device. In this regard, our proposed solution meets all the technological safeguards recognized by Reidenberg and Schaub (2018) for privacy with respect to big data in education, i.e., it implements the necessary technical mechanisms to assure transparency about data collection, processing, and use; accountability for analytics algorithms and algorithmic decision making; and securing and protecting learning analytics data as sensitive data (Liu et al., 2019; Munoz-Arcentales, 2019). We believe that enabling idiographic learning analytics can offer a significant improvement on the way to responsible learning analytics. In fact, by giving the students the responsibility of their learning, idiographic analytics could be the first step towards responsible LA.

REFERENCES

Bergdahl, N., Nouri, J., Karunaratne, T., Afzaal, M., & Saqr, M. (2020). Learning Analytics for Blended Learning - A systematic review of theory, methodology, and ethical considerations.

International Journal of Learning Analytics and Artificial Intelligence for Education, *2*(2), 46–79.

Brickwood, K. J., Watson, G., O'Brien, J., & Williams, A. D. (2019). Consumer-Based Wearable Activity Trackers Increase Physical Activity Participation: Systematic Review and Meta-Analysis. *JMIR mHealth and uHealth*, 7(4), e11819. https://doi.org/10.2196/11819

Cook, C. R., Kilgus, S. P., & Burns, M. K. (2018). Advancing the science and practice of precision education to enhance student outcomes. *Journal of School Psychology*, *66*(December 2017), 4–10. https://doi.org/10.1016/j.jsp.2017.11.004

Epskamp, S., van Borkulo, C. D., van der Veen, D. C., Servaas, M. N., Isvoranu, A. M., Riese, H., & Cramer, A. O. J. (2018). Personalized Network Modeling in Psychopathology: The Importance of Contemporaneous and Temporal Connections. *Clinical Psychological Science*, *6*(3), 416–427. https://doi.org/10.1177/2167702617744325

Hofman, A., Jansen, B., de Mooij, S., Stevenson, C., & van der Maas, H. (2018). A Solution to the Measurement Problem in the Idiographic Approach Using Computer Adaptive Practicing. *Journal of Intelligence*, *6*(1), 14. MDPI AG. Retrieved from http://dx.doi.org/10.3390/jintelligence6010014

Jacob, R., & Parkinson, J. (2015). The Potential for School-Based Interventions That Target Executive Function to Improve Academic Achievement: A Review. *Review of Educational Research*, *85*(4), 512–552. https://doi.org/10.3102/0034654314561338

Liu, S., Chai, H., Liu, Z., Pinkwart, N., Han, X., & Hu, T. (2019). Effects of Proactive Personality and Social Centrality on Learning Performance in SPOCs. *Proceedings of the 11th International Conference on Computer Supported Education*, *2*, 481–487. https://doi.org/10.5220/0007756604810487

Kirkwood, A., & Price, L. (2014). Technology-enhanced learning and teaching in higher education: what is ‘enhanced’ and how do we know? A critical literature review. *Learning, Media and Technology*, *39*(1), 6–36. https://doi.org/10.1080/17439884.2013.770404

Molenaar, P. C. M., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, *18*(2), 112–117. https://doi.org/10.1111/j.1467-8721.2009.01619.x

Munoz-Arcentales, A., López-Pernas, S., Pozo, A., Alonso, Á., Salvachúa, J., & Huecas, G. (2019). An Architecture for Providing Data Usage and Access Control in Data Sharing Ecosystems. *Procedia Computer Science*, 160, 590–597. https://doi.org/10.1016/j.procs.2019.11.042

Reidenberg, J. R., & Schaub, F. (2018). Achieving big data privacy in education. Theory and Research in Education, 16(3), 263–279. https://doi.org/10.1177/1477878518805308

Saqr, M. (2015). Learning analytics and medical education. *International Journal of Health Sciences*, *9*(4).

Saqr, M. (2017). Big data and the emerging ethical challenges. *International Journal of Health Sciences*, *11*(4).

Saqr, M., & Wasson, B. (2020). COVID-19: Lost opportunities and lessons for the future. *International Journal of Health Sciences*, *14*(3).

Saqr, M., & Lopez-Pernas, S. (2021). Idiographic Learning Analytics: A single student (N=1) approach using psychological networks. *Companion Proceedings 11th International Conference on Learning Analytics & Knowledge (LAK21)*, *in-press*.

Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. American Behavioral Scientist, 57(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Tsai, Y. S., Whitelock-Wainwright, A., & Gašević, D. (2020). The privacy paradox and its implications for learning analytics. ACM International Conference Proceeding Series, 230–239. https://doi.org/10.1145/3375462.3375536

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. https://doi.org/ 10.1016/j.chb.2018.07.027

Winne, P. H. (2017). Leveraging Big Data to Help Each Learner and Accelerate Learning Science. Teachers College Record, 119(3).

Wright, A. G. C., Gates, K. M., Arizmendi, C., Lane, S. T., Woods, W. C., & Edershile, E. A. (2019). Focusing Personality Assessment on the Person: Modeling General, Shared, and Person Specific Processes in Personality and Psychopathology. Psychological Assessment, 31(4), 502–515. https://doi.org/10.1037/pas0000617

Zaff, J. F., Donlan, A., Gunning, A., Anderson, S. E., McDermott, E., & Sedaca, M. (2017). Factors that Promote High School Graduation: A Review of Literature. Educational Psychology Review, 29(3), 447–476. https://doi.org/10.1007/s10648-016-9363-5

Towards Ethical Learning Analytics in Learning at Scale

Dilrukshi Gamage
University of Moratuwa, Sri Lanka
Department of Computer Science and Engineering
dilrukshi.gamage@gmail.com

Olga Viberg
KTH Royal Institute of Technology, Sweden
Department of Media Technology and Interaction Design
oviberg@kth.se

ABSTRACT: Learning Analytics (LA) provides tools and techniques that enable researchers to study and benchmark institutions, learners and teachers. In the context of large scale open online courses such as massive online open courses, LA needs careful ethical considerations in making decisions based on analytics. Due to the COVID-19 pandemics, higher demands for online learning are leading to many relevant analytics. These analytics have made it possible and easier to predict future learning outcomes, make more adequate critical decisions, and take more effective actions respectively. However, the growing body of LA research in learning at scale provides few guidelines to related ethical implications. We discuss possible outcomes as there are limited insights into ethical considerations for studies in learning at scale, specifically in the context of open and scalable learning. Based on our experience and prior evidence of data science research, diversity and equity is a key in the data representation and decision made without considering these may impact long term consequences. Our main contribution is to challenge researchers to engage critically with ethical issues when conducting LA in learning at scale and develop their own understanding of ethically appropriate approaches which will not oppress marginalized groups.

Keywords: MOOCs; Ethics; Learning at Scale; Diversity; Equity

1 INTRODUCTION

The Society for Learning Analytics and Research defines learning analytics (LA) as “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environment in which it occurs''. The needs for LA emerged to improve and optimize learning and benchmark the learning environments. LA achieves several benefits, including the prediction of students’ performances, the development of personalized learning experiences and the increase of learner retention rates. However, when it is applied to the education data stream, rising constraints have been identified, including ethical issues and data privacy (Slade & Prinsloo, 2013; Kiito & Knight, 2019; Lieberman, 2020). Other than this, over-granularity of interpretation, misinterpretation of results due to human judgment factors, focus on reporting but not decision may also have contributed to the rising ethical considerations in LA (Papamitsiou, 2014; Baker & Hawn, *under review*).

In contrast to the LA research performed in 'closed' learning environments such as university, the LA conducted in massive online open courses (MOOCs) or open learning environments are facing challenges. LA researchers often encountered with multicultural demography of students' data and their awareness of the cultural behavioral patterns and digital equity, specifically the contextual background data certain learning outcomes or learning behavior are limited. For example, a predictive model that identifies students in the risk level and to offer help significantly biases on gender and raise (Lee & Kizilcec, 2020). On the other hand, the impact that can bring by LA in MOOCs by providing strategies on how courses should be strategically designed to include diversity and inclusion (Kizilcec & Kambhampaty, 2020).

The purpose of this discussion paper is to explore the key ethical issues regarding the use of LA in the direction of learning at scale. Typically, the process of LA is hindering many important areas which LA researchers themselves are not aware of. At the same time, LA process is lacking a thoughtful LA framework where analytics raised questions to its outcomes. Specifically, the quality of obtained data, problems such as incomplete segments or polluted information which will create negative effects to the outcomes or even building assumptions of representation will lead false positive outcomes (see e.g., Baker & Hawn, under review). It is extremely difficult for LA researchers to provide a holistic overview of students in online courses based on just the data left in a platform. LA scholars must be trained to question and make objective assumptions based on missing data, missing representation and missing interactions in their LA pipeline. In this workshop, we aim to contribute to the discussion of the ethical implications of the LA results based on the incomplete data, and the LA process that ultimately raise ethical concerns relating to the LA research in the context of MOOCs.

In the next section, we summarize key areas that challenge the LA to raise ethical considerations. We begin with the general concerns of LA implications and ignite the discussion in the direction to contrast in learning at scale contexts such as MOOCs. Next, we focus on specific areas of concerns in the MOOC's context, in which LA researchers need increased awareness and addition guidance in terms of enabling a responsible approach to the examination of student data. Finally, we discuss measurements needed to in mitigating the LA challenges in MOOCs.

2 ETHICAL CHALLENGES IN LA

2.1 Security

In general, LA ethical implications apply to the security concerns, specifically, in the context where the databases which store the students' records and activities may indicate very private data. The common concerns arise in lieu to the questions such as: How users should be accessing this information without a compromise? This is typically applicable to MOOCs and precautions to secure data are highly encouraged.

2.2 Privacy

LA can reveal personal information about learners. Similarly, MOOCs' datasets may hold sensitive information such as emails, names or addresses. Privacy has been considered as a threat in LA (Papamitsiou, 2014) and as a constraint (Hoel & Chen, 2018). Some solutions have been proposed such as to anonymize the traceable data such as names and emails or encryption, or increasing restrictions (Khalil, Taraghi, & Ebner 2016). Scholars have also proposed principles that could be used to further develop an educational maxim for data privacy in learning analytics including privacy and data protection in LA that could be achieved by negotiating data sharing with each student (Hoel & Chen, 2018).

2.3 Ownership

In typical LA by closed learning environments, it is often explicit who has the authority and ownership of the data. However, questions related to “who owns the analyzed data of MOOCs” always trigger questions with the possibility of web data availability to crawl. Participants like to keep their information confidential, but at the same time, consent policy is essential to ensure transparency. Further, MOOC providers are encouraged to delete or de-identify personal information of their participants. The consent for collecting data, related to ownership of data, is not explicit in many MOOC platforms where it clearly declares the usage of students’ data. Policies with legislation frameworks should include rules of a collection of personal information and a description of information usage, such as research purposes or third-party information selling.

2.4 Transparency

Secret processes can hide unfair decision making when analytics is applied on educational datasets (Sclater, 2014). By the same token, when LA is applied in the context of MOOCs, providers need to disclose their approach to collecting, analyzing and using participants' data. At the same time, a point of balance should be made when the LA algorithms or tools are proprietary. Sclater (2014) argued different code of practices regarding transparency.

3 ETHICAL CHALLENGES: LA & LEARNING AT SCALE

3.1 Sample Size

One of the key concerns in analytics is to rely on the available data sets (Khalil, Taraghi, & Ebner 2016). Often the datasets are limited with limited sample size compared to the population data and the key question is whether the data provide enough representation to the population? This may lead to false positive data or even bring situations with false negative situations hiding the true effects. Researchers discuss situations supposedly if a group of students ‘gaming the system’ and an analyst builds a prediction model for all students based on MOOCs indicators’ fulfillment, then a false positive action is triggered. As a matter of fact, LA is not only based on numbers and statistics.

3.2 Assumptions and Biases

Assumptions are important metrics that LA researchers draw in LA; for example, it is commonly found that MOOCs’ discussion forums’ activity correlates with learners’ performance or social interactions based on forum posts. Some researchers approved that more social activity in forums is reflected positively on performance while others go against this theory. Conclusions are mainly influenced by the assumptions made and biases in these assumptions based on researcher experience (Slade & Prinsloo, 2013). At the same time, the judgements they draw from the data may not be representing the total population. For example, MOOCs’ providers often assume that higher engaging learners are from Global north. However, it is evident that only a small sample of Global south participants enroll in MOOCs, or when they enroll, they are culturally different in engaging online, or some form informal groups to discuss in their own language.

3.3 Flaws in the LA process

Analytics could fail and thus, mistaken interventions or predictions occur. Failures could happen during the main processes of the LA cycle. Wrong actions in collecting data from MOOCs, errors in processing or filtering and mistaken interpretation of data are possible scenarios of fallacy analytics. Additionally, presenting the results through visualizations might also be within the same page. These flaws may not be intentional yet may be accidental. Predictions with such a process will lead to flaws in decisions (Gardenier & Resnik 2002). On the other hand, when LA process a pipeline, the data is often collected in terms of quantitative measurements. This may not provide meaningful interpretation of relationships. LA researchers are less prepared to consume qualitative data which resonate some relationships which could incorporate in the LA decision process.

4 MITIGATING ETHICAL CHALLENGES OF LA IN LEARNING AT SCALE

Ethical consideration can be seen as a cornerstone of any research endeavor and it can be argued that ethics are integral to professional academic practice (Bruhn et al., 2002). The need for increased ethical awareness has been highlighted for research into teaching and learning at scale (Shum & Ferguson 2012; Slade & Prinsloo 2013). However, others (Khalil et al. 2015) claim that there is a surprising dearth of relevant literature on the ethical considerations of research on specific learning technology-MOOCs (the few articles on the subject include Esposito, 2012; Rolfe, 2015; Marshall, 2016). LA researchers often perform analyses in the context of MOOCs (often due to an easy access to big data), and there are some ethical methods and practices discussed in the literature (Ferguson & Buckingham Shum, 2012; Siemens, 2013; Manca, Manca, Caviglione, & Raffaghelli, 2016), specifically Slade & Prinsloo, (2013) propose key area to consider implementing learning analytics ethics, such as: 1) Who benefits and under what conditions? 2) Conditions for consent, de-identification and opting out 3) Vulnerability and harm 4) Consider where Collection, analyses, access to and storage of data take place. Similarly, based on our experience conducting LA for open learning platforms, we would propose following key directions to consider before applying LA to mitigate the ethical concerns in LA.

- Actively seek knowledge of your population sizes; critically question the dataset of sample size, representation and only use LA for the purpose of improving learner effectiveness and teacher support but not to harm either. Actively seek a dataset which represents wider groups of diverse students.

- LA researchers, educators, data collectors, and other stakeholders need data literacy skills, where in case of LA researcher aware the context of data collection, the informed consent from users, and be aware of unconscious biases and assumptions made for analysis.

- Proactively seek for data which has less representation, increase representation.

- Work towards increased equality and justice, expanding awareness of ways in which analytics have the potential to increase or decrease these and understanding of the value, ownership, and control of data.

REFERENCES

Australian Psychological Society. (2008). *Substance abuse: Position statement.* Retrieved from http://www.psychology.org.au/publications/statements/substance

Baker, R., & Hawn, A. (under review). Algorithmic Bias in Education. Pre-print DOI: https://edarxiv.org/pbmvz/ 2021-03-25

Balakrishnan, R. (2006, March). *Why aren't we using 3D user interfaces, and will we ever?* Paper presented at the IEEE Symposium on 3D User Interfaces. http://dx.doi.org/10.1109/vr.2006.148

Gagné, M., Forest, J., Vansteenkiste, M., Crevier-Braud, L., van den Broeck, A., Aspeli, A. K., . . . Westbye, C. (2015). The Multidimensional Work Motivation Scale: Validation evidence in seven languages and nine countries. *European Journal of Work and Organizational Psychology, 24*(2), 178-196. http://dx.doi.org/10.1080/1359432x.2013.877892

Gardenier, J., & Resnik, D. (2002). The Misuse of Statistics: Concepts, Tools, and a Research Agenda. *Accountability in Research* 9(2), 65–74. https://www.tandfonline.com/doi/abs/10.1080/08989620212968.

Hoel, T., & Chen, W. (2018). Privacy and data protection in learning analytics should be motivated by an educational maxim—towards a proposal. *RPTEL 13*, 20. https://doi.org/10.1186/s41039-018-0086-8

Khalil, M., & Ebner, M. (2015). Learning Analytics: Principles and Constraints. *EdMedia + Innovate Learning 2015*(1), pp. 1789–99.

Khalil, M., Taraghi, B., Ebner, M. (2016). "Engaging Learning Analytics in MOOCS: The Good, the Bad, and the Ugly." In International Conference on Education and New Developments 2016, Slovenia. http://arxiv.org/abs/1606.03776 (February 10, 2021).

Kitto, K., & Knight, S. (2019). Practical ethics for building learning analytics. *British Journal of Educational Technology, 50*(6), 2855-2870.

Kizilcec, R., & Kambhampaty, A. (2020). Identifying Course Characteristics Associated with Sociodemographic Variation in Enrolments across 159 Online Courses from 20 Institutions. *Plos One,* https://dx.plos.org/10.1371/journal.pone.0239766

Kizilcec, R., Piech, C., & Schneider, E. (2013). Deconstructing Disengagement: Analyzing Learner Subpopulations in Massive Open Online Courses. In ACM International Conference Proceeding Series, New York, New York, USA: ACM Press, 170–79. http://dl.acm.org/citation.cfm?doid=2460296.2460330

Hansol, L. & Kizilcec, R. (2020). Evaluation of Fairness Trade-Offs in Predicting Student Success. https://arxiv.org/pdf/2007.00088v1.pdf

Lieberman, M. (2020). Massive shift to remote learning prompts big data privacy concerns. EducationWeek. https://www.edweek.org/technology/massive-shift-to-remote-learning-prompts-big-data-privacy-concerns/2020/03

Manca, S., Caviglione, L. & Raffaghelli, J. (2016). Big Data for Social Media Learning Analytics: Potentials and Challenges. *Journal of e-Learning and Knowledge Society, https://www.learntechlib.org/p/173464/*

Papamitsiou, Z., & Economides, A. (2014). Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. *Educational Technology & Society 17*(4): 49–64.

Sclater, N. (2014) *Code of practice for learning analytics: A literature review of the ethical and legal issues*, Retrieved from JISC, Bristol, UK Website: https://repository.jisc.ac.uk/5661/1/Learning_Analytics_A-_Literature_Review.pdf

Buckingham Shum, S. & Ferguson, R. (2012). Social Learning Analytics. In LAK '12: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, New York, New York, USA: International Forum of Educational Technology and Society, 3–26. http://dl.acm.org/citation.cfm?doid=2330601.2330616.

Slade, S. & Prinsloo, P. (2013). Learning Analytics: Ethical Issues and Dilemmas *American Behavioral Scientist 57*(10), 1510–29. http://journals.sagepub.com/doi/10.1177/0002764213479366 (February 10, 2021).

Gagné, M., Forest, J., Vansteenkiste, M., Crevier-Braud, L., van den Broeck, A., Aspeli, A. K., . . . Westbye, C. (2015). The Multidimensional Work Motivation Scale: Validation evidence in seven languages and nine countries. *European Journal of Work and Organizational Psychology, 24*(2), 178-196. http://dx.doi.org/10.1080/1359432x.2013.877892

Wolfe, J. M., Kluender, K. R., Dennis, L. M., Bartoshuk, L. M., Herz, R. S., Lederman, S. J., & Merfeld, D. M. (2015). *Sensation & perception* (4th ed.). Sunderland, MA: Sinauer Associates.

CROSSMMLA Futures: Collecting and analysing multimodal data across the physical and the virtual

Daniel Spikol
University of Copenhagen
ds@di.ku.dk

Xavier Ochoa
New York University
xavier.ochoa@nyu.edu

Marcelo Worsely
Northwestern University
marcelo.worsley@northwestern.edu

Daniele Di Mitri
DIPF
dimitri@dipf.de

Mutlu Cukurova
University College London
m.cukurova@ucl.ac.uk

Roberto Martinez-Maldonado
Monash University
Roberto.MartinezMaldonado@monash.edu

Jan Schneider
DIPF
schneider.jan@dipf.de

ABSTRACT: Workshop proposal for CrossMMLA focused on collecting and analysing multimodal data across the physical and the virtual. Under the current global pandemic, cross physical and virtual spaces play a substantial factor and challenge for MMLA, which is focused on collaborative learning in physical spaces. The workshop proposes an asynchronous format that includes pre-recorded video demonstrations and position papers for discussion, followed by a half-day virtual meeting at LAK'2021.

Keywords: Learning Analytics, Multimodal Learning Analytics, Hybrid Learning Spaces

1 INTRODUCTION

Over the last several years, Multimodal Learning Analytics (MMLA) has brought together diverse fields that combine educational, computational, psychological, and related research into how people learn and how these complex processes are supported with technology. SOLAR's Special Interest Group on Multimodal Learning Analytics Across Spaces (Cross-MMLA SIG) aims to promote research that considers the challenges of making sense of complex educational data that involve multiple interaction modalities, people, and learning spaces. Understanding and optimizing learning traces from the real world requires new degrees of sophistication across technology, learning, and design; and building upon the ongoing and previous work from the Learning Analytics and related communities.

The workshop aims to explore how learning analytics can effectively capture students' learning experiences across diverse learning that include practice-based activities (medical simulations, sports, field-based science, vocational trades). The core challenge is to capture these interactions in a meaningful way that has been translated as part of formative assessment in real-time and post-reflective reviews (Di Mitri et al., 2018; Echeverria et al., 2019). However, under the current global pandemic, the notion of cross physical and virtual spaces plays a substantial factor and challenge for MMLA, which has is focused on collaborative learning in physical spaces.

Overall, the rapid global shift to virtual learning challenges the research and practice for this field and educational practices. MMLA needs to develop theories about the analysis of human behaviours during diverse learning processes across spaces and create practical tools that could augment learners' and instructors' capabilities. These tools and practices need to be designed and implemented in ethical and sustainable ways to provide value and equity for all learners.

The workshop will serve as a forum to exchange ideas on how we as a community can use our knowledge and experiences from CrossMMLA to design new tools to analyse evidence from multimodal and multisystem data. How do we extract meaning from these increasingly fluid and complex data coming from different kinds of transformative learning situations, and how to best feedback these analyses' results to positively support those learning processes?

1.1 Background

MMLA combines the power of affordable sensor technologies and advances in machine learning to observe and analyse learning activities (Blikstein & Worsley, 2016; Ochoa, 2017). This technology acts as a virtual observer and analyst of learning activities across multiple contexts between stakeholders, devices, and resources. Work by current researchers explores how real-time and automatic video and audio analysis can support learning by automating the analysis of these activities through the development of new tools and methods (Chan et al., 2020; Chejara et al., 2020; Kasparova et al., 2020). Martinez-Maldonado and colleagues (Martinez-Maldonado et al., 2020) are carrying work to streamline multimodal data into meaningful layers that explain critical insights to teachers and students. The potential of the approach to create learning analytics interfaces that communicate insights on team performance and concerns in terms of accountability and automated insights discovery. Researchers have also realized that multimodal data collection in the learning sciences demands new and powerful methodological and analytical techniques and technologies. Noroozi and colleagues (Noroozi et al., 2019) highlight the issues for learning scientists to handle, analyse, and interpret complex and often invisible multimodal data when investigating the regulation of learning in collaborative settings. Molenaar and colleagues (2020) are developing tools for learners to use personalized visualizations for Self-Regulated Learning in Adaptive Learning Technologies that highlight the hidden cognitive, social, and emotional aspects of learning.

1.2 Aim of the Workshop

However, the dimensions and contexts of MMLA are complex and layered and provide researchers with multiple challenges. In the current world situation, both research and practice are further complicated by the necessity of remote learning that includes mixed scenarios with virtual co-located and face-to-face learning activities. The MMLA community urgently needs to find ways to research, design, and further develop our tools and methods to investigate across this new landscape.

The workshop aims to discuss the following actively:

- How can MMLA contribute to support hybrid/virtual learning initiatives across physical and digital spaces?

Researchers and education providers have been adapting to local regulations because of the COVID-19 disruption, and education has been re-invented worldwide. Therefore, our workshop's larger aim is to investigate what roles can MMLA as a community have in supporting this adaptation in the short term and how we can joint efforts to prepare ourselves against the next disruption (in the mid-long term).

2 PRE-WORKSHOP ARRANGEMENTS

This workshop continues a recently established but already very consistent tradition of workshops on multimodal learning analytics (MMLA) and across-spaces learning analytics (CrossLAK). These past events have leveraged various formats, from hands-on learning experiences and tutorials, based on participant contributions/papers and conceptual and community-building activities (which have eventually led to the creation of a Special Interest Group within the Society for Learning Analytics Research).

We proposed an asynchronous format that includes pre-recorded video demonstrations and position papers for discussion that allow for an engaging workshop. For video demonstrations and position papers, and an online web platform was provided for viewing two weeks before the workshop. Before the workshop, we will launch a call for submissions that will shape the demonstration part. The submissions for the demonstration may include one or more of the following:

The submissions for the position papers are focused more directly on the theme of this year's CROSSMMLA workshops crossing the physical and digital learning landscape to support learning under pandemic times. A special focus will on how MMLA can contribute to research and practice to support learning along with these themes:

- Logistical (related to the organisation and planning of multimodal data collection, implementations of MMLA in real-world settings, fidelity issues, real-world evaluations etc.)

- Methodological (related to the system of methods, analyses, technical improvements, data cleaning, pre-processing, and other techniques.)
- Ethical (related to the moral principles and aspects of the MMLA work, fairness, transparency, accountability, surveillance, performance-orientation (Cukurova et al., 2020)

2.1 Important Dates

Date	Activity
29 November 2020	Workshop calls for participation announced
10 January 2021	Workshop Papers Deadline
9 February 2021	Camera Ready Deadline for Demo and Position Papers
21 February 2021	Early Bird Registration Ends
31 March 2021	Web Video Platform Live for demo and position papers
12 April 2021	Half-day at LAK'21 Conference Virtual

3 WORKSHOP OBJECTIVES

The workshop was asynchronous with the videos and discussion available to interested parties from late March, allowing participants (interested people) to watch and start conversations about the position papers and demonstrations. Participants will sign-up for the focused panels based on their interests from participation with the online platform. The workshop is planned to occur during the main conference's pre-conference schedule and planned for a half-day format of up to 4 hours (April 11 or 12, 2021). The workshop is divided into four parts.

- Introductions and workshop overview
- Breakout sessions for in-depth discussions about posters and demos
- Sessions overviews and general discussion
- CROSSMMLA SIG and next steps
- Aside from the (intangible, but very important) learning of participants about CrosssMMLA and the strengthening of the SoLAR Special Interest Group (SIG) on CrossMMLA.

Practicalities, we are investigating for supporting the workshop:

- Web-platform for hosting videos and discussions
- Video Conference tool like REMO for easier breakout sessions

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.

Chan, M. C. E., Ochoa, X., & Clarke, D. (2020). Multimodal learning analytics in a laboratory classroom. In Intelligent Systems Reference Library (Vol. 158, pp. 131–156). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-13743-4_8

Chejara, P., Prieto, L. P., Ruiz-Calleja, A., Rodríguez-Triana, M. J., Shankar, S. K., & Kasepalu, R. (2020). Quantifying collaboration quality in face-to-face classroom settings using mmla. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12324 LNCS, 159–166. https://doi.org/10.1007/978-3-030-58157-2_11

Cukurova, M., Giannakos, M. and Martinez-Maldonado, R. (2020), The promise and challenges of multimodal learning analytics. Br J Educ Technol, 51: 1441-1449. https://doi.org/10.1111/bjet.13015

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

Echeverria, V., Martinez-Maldonado, R., & Shum, S. B. (2019). Towards collaboration translucence: Giving meaning to multimodal group data. Conference on Human Factors in Computing Systems - Proceedings, 1–16. https://doi.org/10.1145/3290605.3300269

Kasparova, A., Celiktutan, O., & Cukurova, M. (2020). Inferring Student Engagement in Collaborative Problem Solving from Visual Cues. Companion Publication of the 2020 International Conference on Multimodal Interaction (ICMI '20 Companion). https://doi.org/10.1145/3395035.3425961

Martinez-Maldonado, R., Echeverria, V., Fernandez Nieto, G., & Buckingham Shum, S. (2020, April 21). From Data to Insights: A Layered Storytelling Approach for Multimodal Learning Analytics. Conference on Human Factors in Computing Systems - Proceedings. https://doi.org/10.1145/3313831.3376148

Molenaar, I., Horvers, A., Dijkstra, R., & Baker, R. S. (2020). Personalized visualizations to promote young learners' SRL: The learning path app. ACM International Conference Proceeding Series, 330–339. https://doi.org/10.1145/3375462.3375465

Noroozi, O., Järvelä, S., & Kirschner, P. A. (2019). Multidisciplinary innovations and technologies for facilitation of self-regulated learning. Computers in Human Behavior, 100, 295–297. https://doi.org/10.1016/j.chb.2019.07.020

Ochoa, X. (2017). Handbook of Learning Analytics. In C. Lange, G. Siemens, A. Wise, & D. Gašević (Eds.), Handbook of Learning Analytics (1st ed.). Society for Learning Analytics Research. https://doi.org/10.18608/hla17

Using MMLA to study the link between body and mind

Tetiana Buraha
Goethe University of Frankfurt
tetianabs@gmail.com

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

Daniele Di Mitri
DIPF | Leibniz Institute for Research and Information in Education
dimitri@dipf.de

ABSTRACT: We collected physiological information from 16 participants engaged in cognitive exercises through the Empatica E4 wristband. The sessions were collected using the LearningHub; the collected sensor data include temperature, blood volume pulse, heart rate variability, galvanic skin response, and screen recording from each participant while performing the exercises. In this paper, we present our setup, describe our dataset, and open the discussion for the data analysis.

Keywords: Multimodal Learning Analytics, Physiological Data, Cognitive Performance

1 INTRODUCTION

The human body is composed of highly specialized interconnected systems. This means that the different systems in the body have some influence on each other. It has been observed that in some cases the influence is bidirectional. The study of Blaesi & Wilson (2010) shows how the internally perceived state of a person influences the posture and movements in the body, while also showing how the posture and movements of the body can influence the internally perceived state of a person. For example, it is common to smile when one feels happy, however, as shown in the study Marmolejo-Ramos et al. (2020) engaging the muscles of a smile has a strong influence on the perception and interpretation of different scenarios.

In the case of cognitive functions, it has been observed how states of powerlessness that have a clear influence on the physiology of a person, can undermine executive functions such as reasoning, task flexibility, attention control, and performance (Derakshan & Eysenck, 2009). While test anxiety can lead to a negative impact on academic performance (Cassady & Johnson, 2002), a certain degree of arousal correlates also to positive academic outcomes (Pijeira et al., 2018).

Current research points out the existence of a link between physiology and cognitive performance. However, the mechanisms determining how these two systems influence each other is not clear. To address this gap, we recorded the temperature, blood volume pulse, heart rate variability, and galvanic skin response from participants engaged in different cognitive exercises. The purpose of this workshop paper is to present our study setup, collected dataset, and discuss possibilities for data analysis.

2 METHOD

Participants

A total of 16 participants eight females and eight males follow the study procedure (see section 2.2). Participants were mostly bachelor students or professionals with a bachelors degree. The age of participants ranges from 19 to 33. All participants volunteered to participate in the study, no identifiable personal data was collected from them.

Procedure

The procedure consisted of playing five different games designed to exercise different cognitive capacities such as working memory, fluid intelligence, math problem-solving, attention, and speed. Each of the participants played each of the games a total of three times.

In the working memory game, participants were presented with a matrix of squares. For three seconds the game shows some squares that are highlighted. After the three seconds are over, participants need to select in the matrix the previously highlighted squares. One gaming session consists of 12 tries. The difficulty of the tries increases or decreases based on the participants' performance.

In the fluid intelligence brain, participants are presented with cards. The cards have a one-digit number and a letter. The cards can appear in a top sloth or in a bottom slot. If a card appears on the top sloth and the card shows an even number then participants need to select the "yes" option, if the number is even they need to select "no". Whenever a card appears in the bottom slot and the card shows a vowel, the participants need to select "yes", otherwise they need to select "no". Participants need to make the selections for 60 seconds without making mistakes and as fast as possible to achieve a high score.

The math problem-solving game consists of single arithmetic operations that appear at the top of the screen and slowly move to the bottom. The participant needs to type the result of the operation before it reaches the bottom of the screen. With time, operations start to appear and to fall faster. Whenever an operation reaches the bottom, the participant loses a life. After the loss of the third life, the gaming session is over.

In the attention game, the participant is presented with circles on the screen. For a fraction of a second, each of the circles shows a different number. Once the numbers vanish, participants need to select the circles in an ascending order based on the numbers that were previously shown. A gaming session consists of 10 different tries. Tries become progressively more difficult or easier based on the performance of participants.

In the speed game, participants have to steer a virtual car to evade obstacles that appear in the highway. The car is in constant acceleration unless it hits an obstacle and goes to a full stop. The game finishes after 90 seconds and the average speed of the car is counted as the participant's performance.

Apparatus and Material

While playing the games, participants wore the Empatica E4 wristband[1]. With the use of this band, we were able to collect the temperature, blood volume pulse, heart rate variability, and galvanic skin response from the participants. We also made screen recordings of all the gaming sessions to collect evidence about the performance of participants. We used the *LearningHub* (Schneider et al., 2018) to create synchronized recordings from the data collected with Empatica and screen recordings. Each recording contains 5 JSON files following the MLT format (Schneider et al., 2018) and a screen recording video. To analyze the recordings, we need first to identify different aspects of the participants' performances and connect them to the collected physiological data. To identify these connections, it is necessary to annotate the data. This annotation process can be performed with the Visual Inspection Tool (VIT) (Di Mitri et al., 2019). The annotated sessions then are stored in the expanded MLT format which can be used for further analysis (Di Mitri et al., 2019).

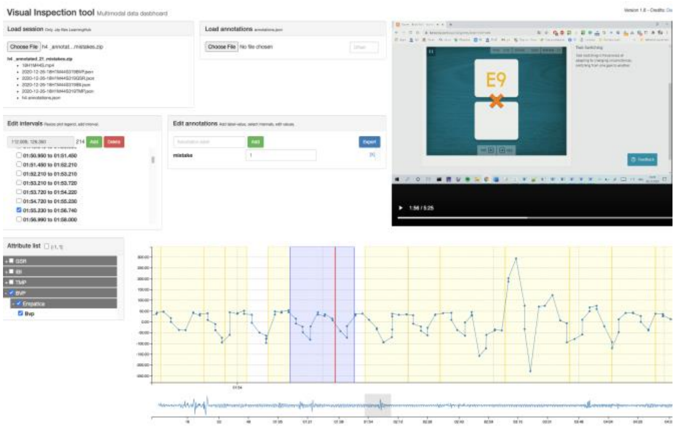


Figure 1: Example of mistake annotated using the Visual Inspection Tool.

3 DISCUSSION

We conducted a preliminary analysis of the dataset for the fluid intelligence game. To study the link between the collected physiological data and the cognitive performance of participants we need to first identify the components that compose the overall performance. For the fluid intelligence game, these components are accuracy and reaction speed.

Our preliminary analysis consisted of classifying the presence or the absence of a mistake comparing the accuracy and F1-score of various machine learning classifiers and using the physiological data as input data. From the six input signals (temperature, blood volume pulse, heart rate variability, galvanic

[1] https://www.empatica.com/research/e4/

skin response and inter-beat interval) we derived more than 3000 features using the tsfresh time-series library (Christ, 2016). These features were reduced to about 100 using Recursive Feature Elimination. We considered 14 recorded sessions containing around 2000 annotated attempts. The unbalanced dataset (9% mistake - 91% not-mistake) was oversampled using the Synthetic Minority Oversampling Technique and validated with 10-fold-cross validation. We trained various supervised models including Naive Bayes and Decision Trees on 13 sessions iteratively leaving one session out for testing. The preliminary results show high accuracy but modest F1 scores. The models we used were time-agnostic. The classification is exclusively based on sensor values in their specific time-interval. Previous or imminent future values (for instance, imminent spikes of BVP due to error -- as shown in Figure 1) are not considered. Further analysis should consider models that account for the temporality of the signals such as Hidden Markov Models or Recurrent Neural Networks. Also, more data will be needed to check the generalisability of the model for unseen participants.

To deepen our understanding of the link between the collected physiological data and the cognitive performance, for future work, we plan to first continue with the analysis of the fluid intelligence game for the speed component of the performance, then the combination of speed and accuracy, and finally analyzing the collected data with the overall performance of participants. Next, the plan is to analyze the dataset for the remaining games.

REFERENCES

Blaesi, S., & Wilson, M. (2010). The mirror reflects both ways: Action influences perception of others. *Brain and cognition*, *72*(2), 306-309.

Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary educational psychology*, *27*(2), 270-295.

Christ, M., Kempa-Liehr, A.W. and Feindt, M. (2016). Distributed and parallel time series feature extraction for industrial big data applications. ArXiv e-prints: 1610.07717 URL: http://adsabs.harvard.edu/abs/2016arXiv161007717C

Derakshan, N., & Eysenck, M. W. (2009). Anxiety, processing efficiency, and cognitive performance: New developments from attentional control theory. *European Psychologist*, *14*(2), 168-176.

Di Mitri, D., Schneider, J., Klemke, R., Specht, M., & Drachsler, H. (2019, March). Read between the lines: An annotation tool for multimodal data for learning. In *Proceedings of the 9th international conference on learning analytics & knowledge* (pp. 51-60).

Marmolejo-Ramos, F., Murata, A., Sasaki, K., Yamada, Y., Ikeda, A., Hinojosa, J. A., Watanabe, K., Parzuchowski, M., Tirado, C., & Ospina, R. (2020). Your face and moves seem happier when I smile. Facial action influences the perception of emotional faces and biological motion stimuli. *Experimental Psychology,* 67 (1), 14-22.

Pijeira-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.

Schneider, J., Di Mitri, D., Limbu, B., & Drachsler, H. (2018, September). Multimodal learning hub: A tool for capturing customizable multimodal learning experiences. In European Conference on Technology Enhanced Learning (pp. 45-58). Springer, Cham.

Reading with and without Background Music: An Exploration with EEG, Eye Movement and Heart Rate

Ying Que[1], Gina M. D'Andrea-Penna[2], Xiao Hu[1], Yueying Dong[2], Andrea A. Chiba[2], John R. Iversen[2]
[1]University of Hong Kong Shenzhen Institute of Research and Innovation
[2]University of California, San Diego
yingque@connect.hku.hk, gdandrea@health.ucsd.edu, xiaoxhu@hku.hk, yud070@ucsd.edu, achiba@ucsd.edu, jiversen@ucsd.edu

ABSTRACT: The effects of background music on learning have been studied in related fields, including psychology and education, but findings are mostly inconclusive. In addition to measurements at the behavioural level, multimodal physiological signals can provide new evidence for exploring the question. This paper presents a pilot study of a reading task for a group of university students whose electroencephalogram (EEG) signals, eye movements, and heart rates were recorded with and without background music. Preliminary results demonstrated the feasibility of integrating multimodal learning analytics to probe the underlying mechanism about the effects of background music on learning.

Keywords: Background Music, Reading Comprehension, EEG Signals, Eye-tracking, Heart Rate

1 INTRODUCTION

Music is a widely-employed stimulus in daily life that elicits entertainment, aesthetic, or spiritual experiences, and/or provides background for other activities such as learning and working. Nonetheless, with regard to the effects of background music on learning, there are many conflicting and inconsistent results. In an recent systematic review (de la Mora Velasco & Hirumi, 2020), some studies have found that background music facilitates memory and recall of information, improves concentration on academic tasks, and enhances mood and emotional states; however, others have reported neutral or negative effects. Two hypotheses, namely arousal-mood-hypothesis and irrelevant-sound-effect, are proposed to explain the mixed results, with focuses on the perspective of emotion or cognition respectively (Li et al., 2020).

Results of existing studies mostly include behavioural measurements such as academic performance and self-reported engagement level, whereas it is known that physiological signals are indicators of cognitive and affective activities (Hu et al., 2019). For instance, electroencephalogram (EEG) signals can reflect the fluctuation of emotional status (Suhaimi et al., 2020). Heart rate variability is deemed as an indicator of cognitive load (Cowley et al., 2013). In terms of eye movements, longer fixation durations and more regressions tend to indicate that the ongoing process of reading is cognitively demanding (Johansson et al., 2012). However, little research has attempted to explore the effects of music on reading with multimodal data input at both behavioural and physiological levels (Hu et al., 2019). New empirical evidence is thus needed to probe the effects of background music on learning and for potentially designing methods that can facilitate the selection of suitable study music.

Recent advancement in learning analytics has begun to examine the cognitive and affective effects of music on learning, such as mental workload and emotional states of students (Hu et al., 2019; Li et al., 2020). This paper presents a pilot user experiment which investigates how background music affects cognitive load and arousal level of students based on analytics of multimodal physiological signals. In particular, this study focuses on reading comprehension, one of the most common learning tasks, with two audio conditions (i.e., background of self-preferred music and silence). In this experiment, participants' interaction logs and multimodal data, including EEG signals, eye movements and heart rates, were recorded simultaneously. Physiological metrics were analysed and compared across audio conditions, revealing interesting results worthy of further exploration.

2 METHODS

2.1 Participants

The pilot study recruited 14 undergraduate students (7 males, 7 females) from a diversified range of majors (e.g. cognitive science, computer science, and bioengineering) in a comprehensive university in the U. S. The mean age of the participants was 22 (SD = 3.0). Five of them reported English as mother tongue. None of them reported visual, hearing or learning impairment.

2.2 Reading Task

Learners acquire knowledge through reading in both physical and digital spaces. In this experiment, participants read eight passages on a computer screen. The passages were selected from GRE-level reading samples, which are generally considered challenging. They covered different topics such as astronomy, geography, history. The experiment contained two blocks, each with two control and two experimental trials. In each trial, participants were tasked to read a passage and answer two questions about its content. The four experimental trials played distinct music pieces in the background from the participants' own choices. In contrast, participants read in silence in the control trials.

2.3 Experimental Apparatus

State-of-art wearable devices were employed to collect multimodal signals. Pupil Core recorded eye movements during reading with sampling rates of 200 Hz in the eye camera and 30 Hz in the world camera. A research-grade wristband, Empatica E4, recorded peripheral physiological signals, including heart rate (HR, 1 Hz), Electrodermal Activity (EDA, 4 Hz), Blood Volume Pulse (BVP, 64 Hz), Skin Temperature (TEMP, 4 Hz). EEG signals were recorded with a 5-channel Cognionics Wireless headset that included electrode sites Cz, Fp1, Fp2, O1 and O2 with a sampling rate of 2000Hz. All of the apparatus could collect data from participants in both physical and digital space. All recorded data were synchronized by timestamps and anonymized to maintain confidentiality.

2.4 Data Analysis

Figure 1 shows the data processing pipeline. Incomplete and invalid data were removed. For each reading period, we removed the first 8 seconds of EEG and heart rate signals, because the data of the first few seconds were likely to be affected by previous activities (Liesefeld, 2018). All of the multimodal signals were segmented on the basis of the start and end time of the reading period corresponding to each passage. Features were further derived from the signal segments according to the corresponding feature extraction methods. After that, we averaged the features in the same audio condition across passages. Finally, paired features were compared between the two audio conditions.

EEG signals were recorded using Lab Streaming Layer (LSL), a middleware for synchronizing signals from multiple sources. High-frequency bands such as alpha, beta and gamma have been taken as effective measures to classify emotions in both valence and arousal dimensions (Suhaimi et al., 2020). Thereby, we used EEGLAB to extract bands of alpha (8-13Hz), beta (13-30Hz) and gamma (30-80Hz). In particular, we computed the mean log spectrum power from the central channel location (i.e., 'Cz').

Eye movements were recorded both through LSL and Pupil Core eye tracker. This paper focuses on eye fixations features which were extracted with Pupil Player, the analytic software coming with Pupil Core. Fixation detection is based on a dispersion-duration method, based on which three measures were calculated. (a) Fixation Number: aggregated fixation counts of each passage. (b) Fixation Duration: aggregated fixation duration of each passage. (c) Mean Fixation Duration: average duration of each fixation.

Heart rates (HR) were recorded using Empatica E4 wristband. Descriptive statistics were extracted from heart rates, including: (a) Mean, (b) Standard deviation (SD), (c) Range.

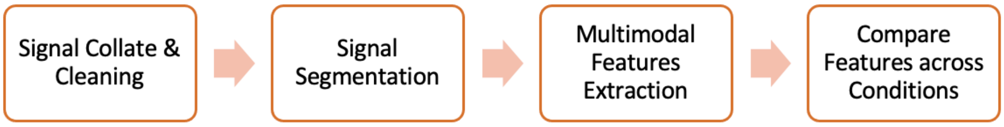
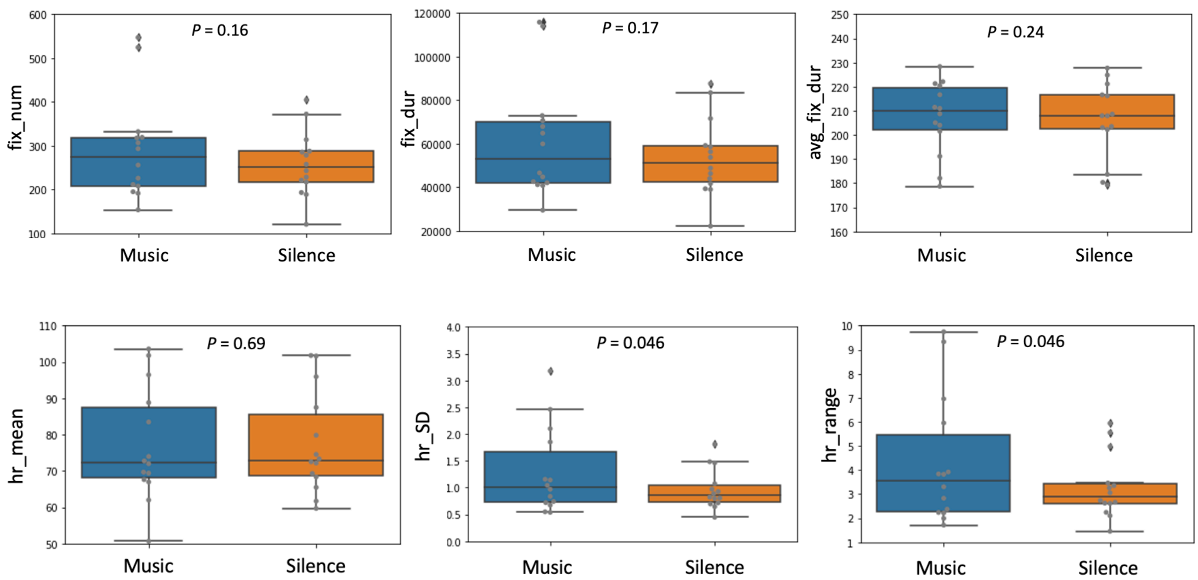


Figure 1. Data Processing Pipeline

3 PRELIMINARY RESULTS

To compare the difference between audio conditions, we first used boxplots to visualize the results from 14 participants (7 males, 7 females) for whom we recorded multimodal signals in terms of eye fixations, HR, and EEG spectra for each audio condition (Figure 2-1, 2-2, 2-3). After that, we applied Paired-sample T-tests to calculate the significant levels of the differences. Preliminary results found significant differences in HR standard deviation and range at p = 0.05 level. No significant differences were detected in measures of HR mean, eye fixation number and duration, average eye fixation duration, or EEG spectra measures at p = 0.05 level.



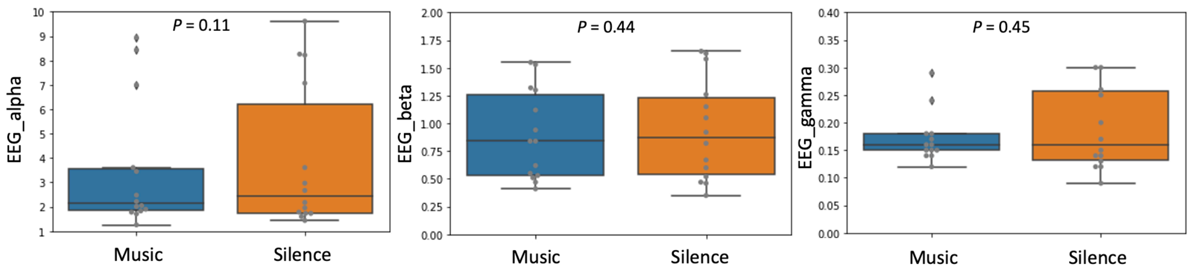


Figure 2-1, 2-2, 2-3. Eye-tracking, HR, EEG Features

4 CONCLUSIONS AND FUTURE WORK

In this study, we reported preliminary results of a within-subject experiment that was conducted with two background audio conditions (i.e., music vs. silence) while participants were engaged in reading comprehension tasks in the digital space. It demonstrated the feasibility of investigating music and learning through a multimodal learning analytics perspective, particularly the mechanisms of collecting multimodal physiological data simultaneously. In the future, we will recruit more participants, analyse fine-grained characteristics of background music selected by the individuals, and interpret the results of multimodal physiological signals from emotional and cognitive perspectives.

ACKNOWLEDGMENTS

This study is supported by National Natural Science Foundation of China (No. 61703357) and the Research Grants Council of the Hong Kong S. A. R., China (No. HKU 17607018).

REFERENCES

Cowley, B., Ravaja, N., & Heikura, T. (2013). Cardiovascular physiology predicts learning effects in a serious game activity. Computers & Education, 60(1), 299-309.

de la Mora Velasco, E., & Hirumi, A. (2020). The effects of background music on learning: a systematic review of literature to guide future research and practice. Educational Technology Research and Development, 68, 2817-2837.

Hu, X., Li, F., & Kong, R. (2019, March). Can Background Music Facilitate Learning? Preliminary Results on Reading Comprehension. In Proceedings of the 9th International Conference on Learning Analytics & Knowledge (pp. 101-105).

Li, F., Hu, X., & Que, Y. (2020, March). Learning with background music: a field experiment. In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge (pp. 224-229).

Liesefeld, H. R. (2018). Estimating the timing of cognitive operations with MEG/EEG latency measures: a primer, a brief tutorial, and an implementation of various methods. Frontiers in neuroscience, 12, 765.

Johansson, R., Holmqvist, K., Mossberg, F., & Lindgren, M. (2012). Eye movements and reading comprehension while listening to preferred and non-preferred study music. *Psychology of music*, *40*(3), 339-356.

Suhaimi, N. S., Mountstephens, J., & Teo, J. (2020). EEG-Based Emotion Recognition: A State-of-the-Art Review of Current Trends and Opportunities. Computational intelligence and neuroscience, 2020.

Immersive Virtual Reality Environment for Training Acute Care Teams

Vitaliy Popov
University of Michigan Medical School and School of Information
vipopov@umich.edu

ABSTRACT: Research on teams in healthcare is that fatal errors due to 'human factors' can occur in 70-80% of medical failures caused by poor communication, ineffective leadership, diagnostic errors, among others. Teams working in high-risk, acute care settings (e.g., trauma, critical care, and emergency medicine teams) are especially prone to these errors. This inquiry provides a unique investigation of medical training using behavioral data from the observation of medical trainees', event stream log data generated by the VR system, and fine-grained "invisible" sensor data about visual attention, emotional arousal, and verbal participation. Multimodal data streams (what do trainees do, attend to, feel and say while a treating cardiac arrest event) will allow us to gain insights into behavioral sequences and interaction patterns that are effective, and which are prone to failure in critical care teams. The unique characteristics of VR simulation environments coupled with the power of multimodal learning analytics and grounded in learning and team sciences theories provide new opportunities to leverage and extend the extant knowledge base about medical teams in ways that have not been possible prior to these new sociotechnical advances.

Keywords: multimodal learning analytics, simulation, virtual reality, team-based training.

1 INTRODUCTION

With in-hospital cardiac arrest survival rates varying between 11% and 35% in the Unites States, patients cared for by clinical staff that has received high-quality resuscitation training have greater odds of survival (Chan et al., 2016). Although resident physicians are often required to have cardiac arrest resuscitation (CAR) training, evidence suggests the need for improved training and assessment methods aimed at increasing patient survival. Simulation-based instruction has been generally accepted as playing an important role in CAR training. Specifically, best practices in healthcare simulation have emphasized the importance of teaching both clinical management and nontechnical (team-based) skills during team training instruction.

Using Virtual Reality (VR) simulations to train effective medical teams has begun to emerge as a viable, innovative and scalable tool in healthcare education (Bracq, Michinov, & Jannin, 2019). When compared to current manikin–based simulation training, VR-based training offers several unique features that provide greater accessibility, collection of dynamic data, improved resource utilization, and increased immersion/realism while still providing opportunities for training healthcare workers on the critical 'high acuity, low frequency' events that are difficult to recreate in real life and allowing them to make decisions, as well as make mistakes, without risk to the patient (Bracq, Michinov & Jannin, 2019). The virtual environment allows for the collection of discrete, nuanced, and dynamic data of team processes that unfold over time that was previously either difficult or impossible to observe during conventional simulation training. The collection of real-time team behavioral data (e.g., behavior, visual attention, speech analysis) can potentially allow for the examination of meaningful associations, observe trends, and provide individualized, learner-specific feedback to each participant. Importantly, whereas conventional simulation provides a more varied experience between learners, VR can offer a more consistent learning experience by providing stimuli that are standardized and responding to learners' responses in a more reliable manner.

1.1 What Makes this Project Innovative or Differentiates It from Current Assessment Methods.

The current standard assessment practices in simulation-based team CAR training are based on the instructor's observation of specific skills and global team performance. The observation tools generally consist of two main approaches: behavioral marker systems and coding schemes (Kolbe & Boos, 2019). These are labor intensive, obtrusive and prone to personal judgment and error. They also tend to result in research that is not replicable or scalable. Most importantly, the current approaches do not capture the process data of how teams dynamically develop over time and how these dynamics can predict team outcomes. Although this assessment methodology aims to be standardized and can provide insights to instructors regarding individual and team performance, many skills such as communication with team members, the use of closed loop communication, verbalizing status changes or changes in team roles, management of distractions, common fixation errors, management of stress and team dynamics as well as many other potential errors are difficult to accurately measured in real time. This diminishes the ability to perform accurate assessments and provide feedback on these high-level skills. A cardiac arrest is an incredibly complex patient care emergency and reliance only on instructor observation and team member self-efficacy can fail to detect a large number of high-level, critical skills that are essential for effective team performance.

To address this gap, this project will use new high-fidelity sensor technology and computational models to capture discrete, nuanced, and real-time data of team processes (e.g., behavior, visual attention, emotional arousal) that unfold over time that was previously either difficult or impossible to observe during conventional simulation training. With the addition of these data, critical feedback can be provided by instructors during simulation debriefing sessions to allow for more targeted intervention and more rapid development of these complex skills. For instance, a sequence of fixations on key AOIs revealing that all team members focusing only on the cardiac monitor instead of patient respirations, or mental status could significantly delay recognition of critical events or potential errors. Similarly, the identification of a single team member with high cognitive load and significant stress response may indicate a low level of familiarity with the specific condition, uncertainty about individual role on the team, or discomfort with team dynamics which may impair overall team performance. The concept of dynamic behavioral tracking has the potential to address several limitations of current assessment methods and unlock the true potential of simulation to train the next generation of residents to provide the best care possible to their patients.

2 METHODS

Four main research questions will guide the development and trial of **iREACT** system (**I**mmersive Virtual **R**eality **E**nvironment for Training **A**cute Care **T**eams) using a cardiac arrest resuscitation scenario to effectively teach both clinical management and teamwork skills. Each phase of the project will be guided by the following research questions:

Phase I: Ethnographic Study, Design of the iREACT prototype and Usability Trials

1. How can multimodal data collection tools (new types of biosensors, computational models and data visualization techniques) be designed and integrated into the VR simulation-based cardiac arrest resuscitation (CAR) training?
2. What are appropriate feedback strategies based on the biosensor data that can help improve clinical knowledge and teamwork skills?

Phase II: Prospective Observational Study

3. In what ways does adding dynamic behavioral tracking to the process measure assessment during simulation-based CAR training support (or hinder) the development of: (a) cognitive skills (situational awareness, decision making); (b) clinical knowledge (aspects of the treatment called for by the mega-code), and/or (c) communication skills?
4. How does feedback provided by faculty based on the quantitative biosensor data collected by iREACT system compare with traditional feedback practices?

2.1 Research setting – Phase I.

We will use existing Health Scholars' VR simulation training on Advanced Cardiovascular Life Support (VR ACLS https://healthscholars.com/acls/) as a case study (see Figure 1). Designed in accordance with the American Heart Association's guidelines and standardized educational programs, VR ACLS training leverages state-of-the-art voice recognition and motion capture technologies to deliver an immersive experience on the management of cardiopulmonary arrest, cardiac arrhythmias, and other cardiovascular emergencies. In this simulation, autonomous agents play the role of the team members and interact with a team leader, the only human in the scene, in a natural manner. Trainees identify cardiac rhythms in the context of the patient's stability and direct virtual team members to shock, give medication, and/or perform CPR as necessary - all under time pressure and rapid workload changes.

Figure 1: Screenshots of a trainee and virtual team members running mega code in the HealthScholar VR ACLS environment.

2.1.1 Design of iREACT system: integrating biosensor modules into Observer XT.

The iREACT system will capture activities and physiological parameters that are specific to individual participants such as pulse rate, galvanic skin response, the direction of their gaze and speech. Measurements of individuals will be done with commercially available and unobtrusive wearables sensors such as the BIOPAC BioNomadix which will record galvanic skin response and pulse rate which are both correlated to stress, as well as other parameters such as 3D acceleration and skin temperature. Learners will also wear pendent microphones to record their specific speech which will be converted to text using services such as Google Cloud Speech and Amazon Transcribe for keyword spotting and speech pattern analysis. Speech recordings will be analyzed to measure the effectiveness of team communication patterns in terms of the amount of speech and who it is directed at, as well as the latency of the responses to the team leader's instructions. Learners will wear HTC VivePro Eye VR headsets equipped with integrated Tobii eye-tracking sensors. These eye tracking sensors operate using the binocular dark pupil tracking technique to measure absolute pupil dilation, gaze direction, gaze origin, and time for each eye fixation.

2.1.2 Multimodal data modelling and analysis

All of these data streams will be collected via an on-site server and the use of the Noldus Observer XT platform (www.noldus.com/observer-xt) for time synchronization, analysis and playback. The Observer XT platform combined with multimodal modeling approaches will allow us to synchronize, segment, and visualize multimodal sets of time-coded information. We will employ Echeverria's modeling representation approach termed the multimodal matrix. To do this, we will create rules based on the Distributed Cognition for Teamwork principles (Rybing et al., 2016), TEAM observation rubric (Cooper et al., 2016) and current American Heart Association algorithms to encode each modality of data (e.g., activity logs, physiological data, speech, gaze direction) into one or more of the columns of a matrix. For example, gaze fixation data is meaningless without a frame of reference (e.g., important areas such as cardiac monitor, patient etc.). Area-of-interest (AOI) analysis, which maps fixations to labeled target areas, will be used to annotate raw gaze fixation data (Salvucci & Goldberg, 2000. These AOIs are meaningful because a learner looks at them to either make a diagnosis or institute a management plan. Rows can then contain information of the visual attention each trainee attends to at every moment, an dthen triangulated with e.g., EDA peaks.

REFERENCES

Bracq, M. S., Michinov, E., & Jannin, P. (2019). Virtual reality simulation in nontechnical skills training for healthcare professionals: a systematic review. *Simulation in Healthcare*, *14*(3), 188-194.

Chan, P. S., Krein, S. L., Tang, F., Iwashyna, T. J., Harrod, M., Kennedy, M., & Nallamothu, B. K. (2016). Resuscitation practices associated with survival after in-hospital cardiac arrest: a nationwide survey. *JAMA cardiology, 1*(2), 189-197.

Cooper, S., Cant, R., Connell, C., Sims, L., Porter, J. E., Symmons, M., Nestel, D., & Liaw, S. Y. (2016). Measuring teamwork performance: Validity testing of the Team Emergency Assessment Measure (TEAM) with clinical resuscitation teams. *Resuscitation, 101*, 97-101.

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* (pp. 1-16).

Kolbe, M., & Boos, M. (2019). Laborious but elaborate: The benefits of really studying team dynamics. *Frontiers in psychology, 10*, 1478.

Rybing, J., Nilsson, H., Jonson, C. O., & Bang, M. (2016). Studying distributed cognition of simulation-based team training with DiCoT. *Ergonomics*, *59*(3), 423-434.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications* (pp. 71-78).

Bridging the Gap Between Theory and Tool: A Pragmatic Framework for Multimodal Collaboration Feedback

Maurice Boothe Jr.*, Collin Yu*, Xavier Ochoa
New York University
{mab1488, cfy229, xavier.ochoa}@nyu.edu
* These authors contributed equally

ABSTRACT: Providing feedback for collaboration activities is divided between theoretical studies of individual constructs and empirical tools that provide feedback. This paper proposes a pragmatic framework that strives to bridge this divide by providing a set of comprehensive and comprehensible constructs that can be estimated through multimodal learning analytics and used to provide actionable feedback to students.

Keywords: multimodal learning analytics, feedback, collaboration framework

1 INTRODUCTION

Collaboration is a topic that has been regularly investigated in traditional educational research. Most of the studies conducted in this field have focused on the definition, validation and/or effect of a given construct in the context of collaboration (e.g., rapport in Gratch, Wang, Gerten, Fast, & Duffy, 2007). On one hand, these studies usually focus on a limited set of constructs since their objective is not to provide feedback to students or instructors, but to build theoretical understanding of the collaboration process. On the other hand, there are approaches for Multimodal Learning Analytics that focus on providing feedback to students based on simple features extracted from the collaborative process, but they are mainly focused on participation and attention without connection to high-level collaboration constructs (Praharaj, Scheffel, Drachsler, & Specht, 2018). The first set of studies produces valuable theoretical knowledge about important aspects of collaboration, but that information doesn't reach the actual participants. The second set provides feedback to participants but is based on simple and disconnected features that do not provide students with a holistic and theory-grounded view of their collaboration process. The future of multimodal tools in providing feedback on collaboration lies in the middle ground. Our work intends to bridge this gap by providing a pragmatic collaboration framework that supports the creation of tools for developing collaboration skills in students while also providing an opportunity to conduct collaboration research.

2 PRAGMATIC COLLABORATION FRAMEWORK

The core of this framework is the set of high-level collaboration constructs that it defines. To be useful, these constructs must be (1) **Fundamental**, representing characteristics that are required for successful collaboration; (2) **Learner-centered**, described in terms of the behavior of the individual rather than the more complex behavior of the group; (3) **Actionable**, including only aspects that can be purposefully practiced and changed by the individual; (4) **Comprehensible**, intuitively relatable to students' collaboration experiences; (5) **Computable**, operationalizable through mapping onto observable behaviors and low-level multimodal features; and (6) **Theory-grounded**, mappable to theoretical and psychometric frameworks. This pragmatic framework will not only facilitate the construction of collaboration feedback tools, but it can also be used to identify gaps between what can be measured currently and what will need to be measured in the future.

We draw upon Hilgard's theory of mind (1980) as the foundation for our framework's constructs. This theory divides the human psyche into three cooperating aspects: cognition (our ability for rational thought); meta-cognition/conation (purposefully striving towards valued goals); and affect (our feelings and emotions). In collaboration, we use these three aspects to interact with others in the group in a bidirectional way between the individual and the group. Each aspect of our framework gives origin to two constructs, one representing the action of the individual towards the group, and the other for the actions that the group provokes in the individual. These constructs were obtained by synthesizing constructs used in human-oriented frameworks for collaboration assessment (Lai, DiCerbo & Foltz, 2017).

2.1 Cognitive Aspect

Contribution (Individual-to-Group): A cognitive action that advances the fulfillment of the collaborative goal, output, or result. Though it seems external in behavior, a *contribution* is considered a cognitive action due to underlying mental efforts required to initiate the observed behavior. Examples of *contribution* indicators include suggesting an idea, providing research, completing an assigned task, making an individual effort towards the solution of a problem, furthering a discussion, or sharing a point of view.

Assimilation (Group-to-Individual): An action that a collaborator takes as they receive a cognitive contribution from their fellow group mates, making sense of a *contribution* and incorporating it into their own mental models. Directly counter to *contribution*, *assimilation*'s directionality comes from the individual receiving ideas from the group, expressed through clarification, or adding to a future contribution. Examples of *assimilation* indicators include asking for clarification on an idea, asking follow-up questions, and mentioning previous ideas as part of a new *contribution*.

2.2 Metacognitive Aspect

Team Coordination (Individual-to-Group): Any action taken to organize the individual efforts of collaborators for the sake of improving the efficiency of their collective effort. An interesting consequence of its communal nature is that any subgroup of the whole can enact *team coordination* upon any other subgroup. Examples of *team coordination* indicators include scheduling meetings, delegating asynchronous tasks, and managing logistics in other ways.

Self-Regulation (Group-to-Individual): The individual cognitive actions through which a group member adapts their behavior to the group and its qualities to facilitate their own participation. While not as readily measurable as other constructs, *self-regulation* remains an important consideration as it relates to the ways in which a collaborator helps him or herself to be effective. Examples of *self-regulation* indicators include maintaining a calendar, taking a break when tired or frustrated, effectively managing a task list, and promptly replying to emails.

2.3 Affective Aspect

Cultivation of Environment (Individual-to-Group): Any action that a group member takes to support or edify another group member. Ideally, these actions encourage other group members and instill in them a desire to take part in and collaborate with the team. These actions can be found to originate from any subset of group members and are directed towards any other subset of group members. Examples of *cultivation of environment* indicators include verbal and non-verbal signals of acceptance (or rejection) with specific support phrases, invitations to participate, and paying compliments.

Integration (Group-to-Individual): The internal decision-action of a group member to read the collaboration environment, take membership, and further bring themselves into the group. In one sense, this could be represented by the “ownership” that one feels towards the group’s actions and products. The action of integration is a combination of the group member’s choice to engage and the state of the group’s current cultivated environment. Examples of *integration* indicators include showcasing the collaboration work in other contexts, intervening to avoid conflict and being physically close.

3 APPLYING THE FRAMEWORK

This pragmatic framework draws from the work of Echeverría (2020, Chapter 4) in mapping high level constructs to multimodal features. First, the six identified constructs for collaboration are mapped to a set of behavioral indicators, human-observable actions that can be used to measure the construct of interest. These behavioral indicators are then mapped to one or more objectively measurable indicators. These second level indicators are unambiguous, machine-calculable measures that assert the presence or absence of the behavioral indicator through the multi-level fusion of multimodal features (Worsley & Ochoa, 2020). Finally, all the low-level multimodal features needed to calculate the objectively measurable indicators are specified. Given the level of detail required, this framework must be applied to each unique collaboration scenario. To illustrate its use, we present applied examples for two specific contexts. In Table 1, the blue cells show the framework applied to a face-to-face activity in a middle school history class in which the students are tasked with the creation of a group presentation and given time to meet synchronously. Orange represents an online, asynchronous collaborative activity conducted in a university database design course in which students collaborate virtually through an online tool that has a common drawing interface and a section for comments and discussion.

4 CONCLUSION

This work is the initial presentation of a pragmatic framework that facilitates the construction of multimodal collaboration feedback tools that are (1) comprehensive, comprehensible, and actionable for students, and (2) grounded in theoretical constructs. Future work for this framework includes validation in participatory workshops involving students and instructors, further mapping to existing collaboration assessment models, and implementation in a real multimodal collaboration feedback tool. However, in its current form, its objective is to spark discussion in the Learning Analytics and Multimodal Learning Analytics communities about the need for such a framework and lay a foundation for bridging the gap between theory and tools.

Construct	Behavioral Indicators	Objectively Measurable Indicators	Multimodal Features
Contribution	Expressing new ideas	New concepts detected in the speech are assigned to the individual that says them.	Speech content Speaker identification
	Establishing new connections between elements	New connections assigned in the Entity-Relationship diagram	Entity-Relationship tool actions
Assimilation	Student paying attention to the contribution of others	Gaze directed to current speaker	Gaze detection Speaker identification
	Student taking into account the ideas of other students	Student linked (or suggested linking) entities created by other students	Entity-Relationship tool actions Comment content
Team Coordination	Assigning roles to group members	Differences measured between speech activities and contributions made by the students	Speech content Position and movement
	Planning logistics of asynchronous work	Tasks assigned to other individuals	Comment content
Self-Regulation	Allowing fair contribution of oneself and others	Speaking time kept equitable relative to other group members	Speech duration
	Compromising on conflicts	Conflict over diagram content negotiated successfully	Entity-Relationship tool actions Comment content
Cultivation of Environment	Supporting the contributions of other group members	Compliments made towards the contribution of others	Speech content Speaker identification
	Catering the workspace to the needs of the team	Jokes contributed within the discussion	Comment content Final survey
Integration	Defusing possible conflicts	Conflict strategically mediated	Speech content Movement and gestures
	Expressing ownership of the project	Project added to portfolio	Student portfolio logs

Table 1. Examples of the mapping the collaboration framework to two contexts. Blue is the K-12 face-to-face scenario; orange is the HE asynchronous scenario.

REFERENCES

Echeverria Barzola, V. I. (2020). Designing Feedback for Collocated Teams using Multimodal Learning Analytics (Doctoral dissertation).

Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. (2007, September). Creating rapport with virtual agents. In International workshop on intelligent virtual agents (pp. 125-138). Springer.

Hilgard, E. R. (1980). The trilogy of mind: Cognition, affection, and conation. Journal of the History of the Behavioral Sciences, 16(2), 107-117.

Lai, E., DiCerbo, K., & Foltz, P. (2017). Skills for Today: What We Know about Teaching and Assessing Collaboration. Pearson.

Praharaj, S., Scheffel, M., Drachsler, H., & Specht, M. (2018, September). Multimodal analytics for real-time feedback in co-located collaboration. In ECTEL (pp. 187-201). Springer, Cham.

Worsley, M., & Ochoa, X. (2020). Towards collaboration literacy development through multimodal learning analytics. In CEUR Workshop Proceedings (Vol. 2610, pp. 53-63). CEUR-WS.

Combining multimodal data to explore emotion during learning with an ALT

Anne Horvers*, Rick Dijkstra, Ard Lazonder, Tibor Bosse & Inge Molenaar
Behavioural Science Institute, Radboud University Nijmegen
*a.horvers@bsi.ru.nl

ABSTRACT: During the ongoing COVID-19 pandemic, students' learning shifted from the physical to the virtual space. Students use adaptive learning technologies and teachers can follow their students' learning progress from a distance via dashboards. While we know the important role of emotion in learning, teachers have less insight into students' emotional states than normally in a physical classroom. This project aims to get a better understanding of the role of children's emotion during learning with an adaptive learning technology. Using a multimodal approach, the objective is to gain insight into the association between physiological arousal, self-reported valence, and observed emotion type of grade five students. The ultimate goal is to incorporate emotion into a teacher dashboard and investigate how teachers use this dashboard to adjust their instruction to students' needs.

Keywords: Emotion, Multimodal data, Adaptive Learning Technologies

1 BACKGROUND

During the COVID-19 pandemic, distance learning became more important and learning shifted to the virtual space. In this virtual space, it becomes harder for teachers to follow their students' learning and adjust instruction to their needs (Vlachopoulos, 2020). When students work with learning technologies remotely, teachers can gain insight into the learning progress of their students using the provided dashboard. However, when learning takes place in the physical space, teachers not only adapt to the students' cognitive needs but also take into account the emotional state of students. Teachers adjust their actions to students' needs using information about their emotions. However when students work from home, the teacher does not have concurrent insight in students' emotions while they are solving problems.

Adaptive learning technologies (ALTs) are suitable to provide students with the appropriate learning materials from a distance (Aleven et al., 2016). These technologies automatically select the next problem a student needs to solve. The difficulty of the problems provided to the student is adjusted based on their ability using an ELO-algorithm (Klinkenberg et al., 2011). However, students' emotions are not visible in the virtual space at this moment. There is an opportunity for ALTs to support teachers in this matter. By providing teachers with a dashboard with information about students' emotions, they get the opportunity to gain more insight into their learning process. Based on the visible information, teachers can adjust their instruction to the needs of students. Before this dashboard can be developed, we need to get a better insight in the role of emotion during learning. This paper proposes a study with a new method of gathering multimodal data about emotions during learning. Previous research shows that emotion plays an important role in learning. Positive emotions, such as enjoyment or pride are positively related to learning, whereas negative emotions such as frustration negatively influence learning (Loderer et al., 2018). Emotion has no generally accepted definition, but many theories and frameworks have been developed. This project builds on the dimensional perspective of emotion, where arousal, valence and object focus are distinguished (Pekrun, 2006). Arousal refers to the physiological activation in the body that occurs when an emotion is triggered. Valence categorizes emotions as either positive/pleasant or negative/unpleasant. Object focus can be focused on ongoing learning activities or on learning outcomes either prospectively or retrospectively.

1.1 Multimodal measurement of emotion

Often three types of emotional responses to personally meaningful stimuli are distinguished; physiological, experiential, and behavioural responses (Mauss & Robinson, 2009). Physiological responses contain the reaction of the body when an emotion is evoked. Experiential responses refer to the subjective personal experience of an emotion and behavioural responses to the visible behavioural response. These different responses can be measured using different modalities. First, physiological responses can be measured by students' heart rate variability (HRV), electrodermal activity (EDA), Blood Volume Pulse (BVP), and skin temperature. Second, experiential responses can be investigated by self-reports. Third, behavioral responses can be observed looking at the facial expressions and body posture of the participants.

However, each specific measurement measures one type of emotional response and lacks the ability to properly measure the other two responses. For example, the physiological response is measured successfully with physiological measurements such as electrodermal activity, but experiential and behavioural responses are not properly measured. Electrodermal activity (EDA) holds a promise for measuring arousal during learning (physiological response), self-reports are used to assess valence (experiential response) and observations are used to determine the type of emotion people show (behavioural response) (Baker et al., 2010; Malmberg et al., 2019; Putwain et al., 2020). Hence, it is worthwhile to combine these three types of measurements. Using multimodal data helps to overcome constraints of a single data stream and enables measurement of all responses to emotions during learning. By combining the three types of emotional responses (physiological, experiential and behavioural), this project aims to get a better understanding of students' emotions during learning. The aim is to give meaning to the peaks in arousal by connecting them to valence and emotion type. By using the combination of these multimodal data streams as a starting point, we will work towards creating a dashboard which provides teachers with information about students' emotions.

2 METHODS: MEASURING EMOTION

Participants are grade five students in Dutch primary education. These students work on arithmetic problems about fractions within an existing adaptive learning technology. To support the researcher during this study and secure data synchronization retrospectively, a real-time dashboard is developed which combines the different multimodal data streams. The Emotion Dashboard shows the correctness of answers to the problems students solve in the ALT as well as the real-time physiological arousal. When the student submits an answer to the ALT, a valence pop-up is prompted to the student to register the students' self-report. The researcher is prompted at the same time to indicate the emotion type by observing the student. The architecture of this dashboard is visible in Figure 1. The dashboard synchronizes the electrodermal activity, valence self-reports and observed emotion type to a csv file. The correctness of the students' answers is also captured within this file.

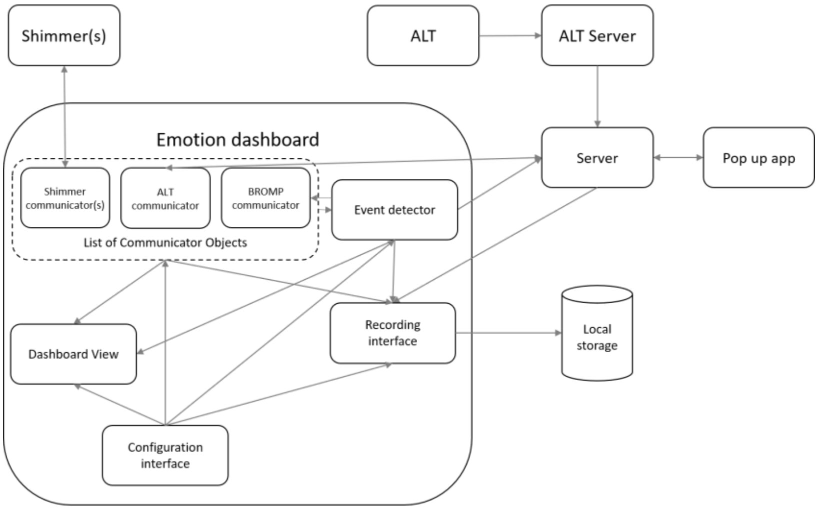


Figure 1: Emotion Dashboard architecture

2.1 Measurements

Answers. The ALT logs the correctness of students' answers and sends them to the emotion dashboard via a server. To be able to synchronize the answer data with the other data a timestamp with millisecond precision is added when the data is received by the dashboard.

Physiological arousal. Physiological arousal (physiological response) is measured by collecting real-time electrodermal activity (EDA data) from Shimmer3 GSR+ wristband. EDA shows the variation in electrical properties of the skin based on sweat gland activity. The Shimmer wristband is placed on the non-dominant hand, with two electrodes placed on the fingers. EDA is measured in micro Siemens at a frequency of 51.2 hertz. Timestamps with millisecond precision are added to each measurement of EDA.

Valence. The valence of the emotion (experiential response) is measured using the Smileyometer (Read, 2008). This is a 5 point-scale with emoticons depicting very negative, negative, neutral, positive and very positive emotions. Timestamps with millisecond precision are added based on the answer data timestamp.

Emotion type. A derivative of the Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) is used to observe the emotion type students show (behavioural response) (Ocumpaugh et al., 2015). First, on-task and off-task behavior is scored. Then in case of on-task behavior, the trained research assistant will select the observed emotion type: enjoyment, boredom, confusion, engaged concentration, surprise, relief, disappointment and frustration. Timestamps with millisecond precision are added based on the answer data timestamp.

3 RELEVANCE AND CHALLENGES

While learning in a virtual space has become more common during the ongoing pandemic, teachers are unable to see how their students emotions when they are working at home. Using the already existing adaptive learning technologies, teachers only have insight into how students are performing on a cognitive level. This project will investigate which role emotions play while students work with ALTs and ultimately create an emotion sensitive algorithm which uses students' emotional information to adjust the difficulty of problems to their needs. The combination of arousal, valence and emotion type with three different modalities (respectively physiological measurements, self-reports and observation) has a high potential to investigate the role of emotion. In the current set-up, valence self-reports and emotion type observations are prompted after every exercise. However, the aim of connecting valence and emotion type to arousal peaks requests real-time peak detection. Previous research has used peak detection in arousal data in hindsight (Boucsein et al., 2012). Currently, we are analyzing arousal data retrospectively using the Ledalab Matlab toolbox. The challenge is to develop a concurrent way to detect peaks while students are learning. Once we have developed such a concurrent peak detection and understand how arousal is related to valence and emotion type, we can inform teachers about students' emotional states when distance learning, for example during the current pandemic.

REFERENCES

Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2016). Help Helps, but only so Much: Research on Help Seeking with Intelligent Tutoring Systems. International Journal of Artificial Intelligence in Education, 26(1), 205–223. https://doi.org/10.1007/s40593-015-0089-1

Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. International Journal of Human Computer Studies, 68(4), 223–241.

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. Psychophysiology, 49(8), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

Klinkenberg, S., Straatemeier, M., & Van Der Maas, H. L. J. (2011). Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation. Computers and Education, 57(2), 1813–1824.

Loderer, K., Pekrun, R., & Lester, J. C. (2018). Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. Learning and Instruction, 70(November 2018), 101162. https://doi.org/10.1016/j.learninstruc.2018.08.002

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 CSCL, 14(4), 467–490.

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. Cognition and Emotion, 23(2), 1–23. https://doi.org/10.1080/02699930802204677.

Ocumpaugh, J., Baker, R. S., & Rodrigo, M. T. (2015). Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) 2.0 Technical and Training Manual. 1–51. https://doi.org/10.1109/EUROCON.2015.7313780

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. https://doi.org/10.1007/s10648-006-9029-9

Putwain, D. W., Schmitz, E. A., Wood, P., & Pekrun, R. (2020). The role of achievement emotions in primary school mathematics: Control–value antecedents and achievement outcomes. British Journal of Educational Psychology. https://doi.org/10.1111/bjep.12367

Read, J. C. (2008). Validating the Fun Toolkit: An instrument for measuring children's opinions of technology. Cognition, Technology and Work, 10(2), 119–128.

Vlachopoulos, D. (2020). Covid-19: Threat or opportunity for online education? Higher Learning Research Communications, 10(1), 16–19. https://doi.org/10.18870/hlrc.v10i1.1179

CoTrack2: A Tool to Track Collaboration Across Physical and Digital Spaces with Real Time Activity Visualization

Pankaj Chejara, Luis P. Prieto, María Jesús Rodríguez-Triana, Shashi Kant Shankar, Reet Kasepalu
Tallinn University
pankajch@tlu.ee, lprisan@tlu.ee, mjrt@tlu.ee, shashik@tlu.ee, reetkase@tlu.ee

Adolfo Ruiz-Calleja
University of Valladolid
adolfo@gsic.uva.es

ABSTRACT: Multimodal learning analytics (MMLA) offers a holistic view on collaboration by going beyond traditional log data collection and taking into account data from the physical space. It holds the potential to support teachers and students during collaborative learning. In this direction, we present a web-based MMLA prototype: CoTrack2. This prototype is the updated version of CoTrack [1] and it also supports online collaboration activity and real time activity monitoring. Teachers can create collaborative learning sessions with the help of its web interface. Students can join these sessions and use Etherpad (a collaborative text editor). It allows students to construct a joint document as the output of their collaborative activity while speaking to each other through an audio/video channel. The tool tracks each of the students' activities (i.e. logs) in Etherpad. The tool also records audio-video data from collaboration sessions. This data can be later utilized for annotation purposes or a detailed understanding of collaboration behavior. CoTrack2 uses a Javascript library[1] to synchronize the clocks between clients and servers. CoTrack2 also provides a dashboard for teachers. This dashboard is updated for every 5 seconds time window. This dashboard has two levels of visualization- group and individual. The dashboard visualizes total activity in the Etherpad and students 'who-is-talking-after-whom' network. CoTrack2 uses voice activity detection to get students' speaking time and speaking sequence. This sequence is then utilized to generate the network. The dashboard also presents collaborative writing information at the individual level (e.g., number of characters added or deleted). The dashboard also provides a view of students' written text with time navigation to inspect the evolution and contributions to the joint document.

Demonstration Movie: https://youtu.be/IOH4S2doZTA

Keywords: Collocated collaboration, Multimodal Learning Analytics, Computer-Supported Collaborative Learning

References

1. Chejara, P., Prieto, L. P., Rodríguez-Triana, M., Ruiz-Calleja, A., & Shankar, S. K. (2020). Cotrack: A tool for tracking collaboration across physical and digital spaces in collocated blended settings. In *Companion Proceedings of the 10th International Conference on Learning Analytics & Knowledge*.

[1] https://github.com/NodeGuy/server-date

Introducing MBOX

Daniel Spikol
University of Copenhagen
ds@di.ku.dk

Hamza Ouhaichi
Malmö University
hamza.ouhaichi@mau.se

Bahtijar Vogel
Malmö University
bahtijar.vogel@mau.se

ABSTRACT: MBOX is a proof-of-concept Multimodal Learning Analytics Internet of Things (IoT) system with multiple sensors that collect data on how small groups of people interact when performing collaborative tasks. These tasks include learning, organizational activities, and engineering tasks (physical computing). MBOX system comprises camera sensors, audio microphone arrays, biometric signals (EEG, HRV, and EDA) that connect to small single-board computers that process data on the edge and stream the meta-data to the cloud. The project aims to help support how people collaborate by providing feedback on physical interactions (body positions, gaze directions, hand motions), voice diarization (amount of talking for each person), physiological feedback, and affective measures (emotional qualities of face and voice).

Keywords: multimodal learning analytics, IoT, edge computing, fog, multi-layered architecture

1 MBOX

This paper presents MBOX, an IoT-based system for capturing multimodal learning analytics data with lightweight systems. The aim is to move away from a centralized system to an IoT approach that allows different sensors to be deployed depending on the learning scenario and computational resources needed. We utilize a multi-layered architecture following the edge-cloud pattern (Portelli and Anagnostopoulos, 2017). This approach integrates all possible computing layers, including Cloud, Fog, and Edge platforms. It has two main advantages: 1) being scalable to collect data from supplementary physical and digital data sources, which is very important for continuous improvement and future evolution of learning settings. Moreover, 2) supporting a de-centralized approach building up from the IoT systems. With MBOX the aim is to promote an adaptation to different learning environments and enable a better scaling of computational resources used within the learning context.

The resulting Architecture for MBOX (see Fig. 1) is edge-fog-cloud-driven approach. The edge part comprises small computation units (single board computers and microcontrollers) and data entry points being sensors, wide-angle camera, microphone array, and biosensors (HRV, EEG, EDA). We are using Timeflux (Clisson et al., 2019), an open-source framework for the acquisition and real-time processing of biosignals that send time-series data to InfluxData[1] that is currently visualized with Grafana[2] hosted in Google Cloud Computing Services[3]. The initial minimum viable prototypes have been deployed. We are now working on the next step, the signal synchronization and data fusion, to investigate the different collaboration patterns (see figure 1).

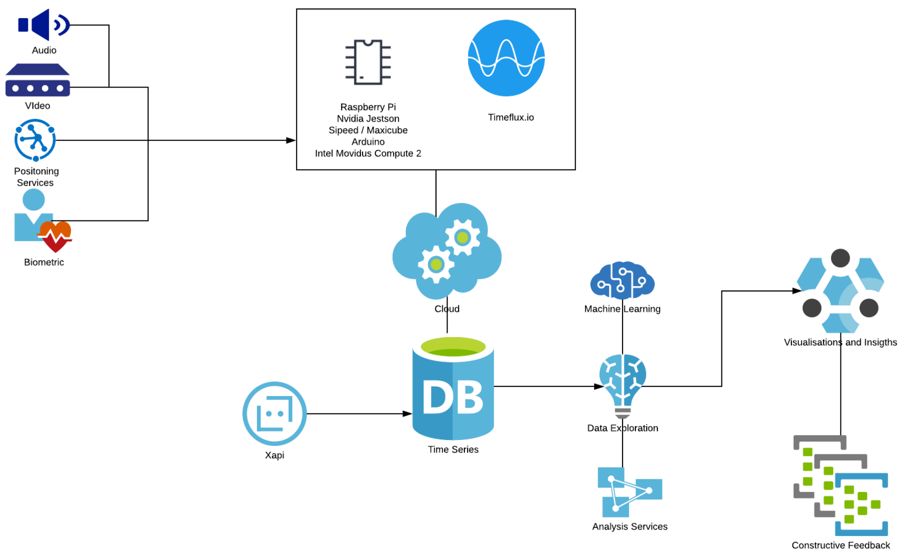


Figure 1: Conceptual Architecture of MBOX

2 VIDEO DEMOSTRATION

The video demonstration will illustrate some parts of the proposed approach with some basic working aspects of the different sensors and the streaming of meta-data to the cloud. Additionally, various edge services that provide essential computer vision and audio operations.

REFERENCES

Clisson, P., Bertrand-Lalo, R., Congedo, M., Victor-Thomas, G., & Chatel-Goldman, J. (2019, September). Timeflux: an open-source framework for the acquisition and near real-time processing of signal streams. In *8th Graz Brain-Computer Interface Conference 2019*.

Portelli and C. Anagnostopoulos, "Leveraging edge computing through collaborative machine learning," in *Proceedings - 2017 5th International Conference on Future Internet of Things and Cloud Workshops, W-FiCloud 2017*, 2017

[1] https://www.influxdata.com/

[2] https://grafana.com/

[3] https://cloud.google.com/

The 7th LAKathon: Is Learning in Isolation a Mission Impossible?

Daniele Di Mitri
DIPF | Leibniz Institute for Research and Information in Education
dimitri@dipf.de

Alan Mark Berg
ICTS, University of Amsterdam
a.m.berg@uva.nl

Gábor Kismihók
TIB Hannover
gabor.kismihok@tib.eu

Jose Ruipérez-Valiente
University of Murcia, Information and Communications Engineering
jruiperez@um.es

Kirsty Kitto
University of Technology, Sydney
kirsty.kitto@uts.edu.au

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

Atezaz Ahmad
DIPF | Leibniz Institute for Research and Information in Education
ahmad@dipf.de

Stefan T. Mol
University of Amsterdam
s.t.mol@uva.nl

ABSTRACT: Welcome to the seventh Learning Analytics Hackathon (LAKathon). The LAKhathon 2021 will become an online laboratory to envisage future Learning Analytics (LA) applications with an emphasis on supporting Learning and mental health through online strategies. Do you have a research question, a dataset or online support orientated idea you would like to explore? Bring it to the LAKathon! We encourage joining this inclusive online 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, Learning Analytics, Online Collaboration, Infrastructure, Mental Health

1 INTRODUCTION

For the last six years, researchers and practitioners have run on-site 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 that have radiated back to the LA research community as a whole. In 2020, with the spread of the Covid-19 pandemic, the attendance of the academic conferences including the LAK conference had to become virtual. This event led to the cancellation of the LAKathon'20. For LAKathon 2021, we have reconsidered the whole process of the LAKathon with the aim of organising it well as the first-ever online LAKathon. All the LAKathons events have been designed to be:

(1) *solution-driven*: participants solve a series of realistic challenges using agile approaches, including brainstorming, design thinking, or fast-prototyping;

(2) *multi-disciplinary*: reflected in the diversity of the participants;
(3) self-organised: we engage in bottom-up, and actionable research questions. Through a Call for Proposals, we aim to elicit research questions that address the yearly topic of "*Is Learning in Isolation a Mission Impossible?*"
(4) *evidence-based*: all LAKathons challenges started from concrete problems, datasets or tools. This is a particularly relevant aspect, as online education generates much more data than the traditional equivalent.

At LAKathon 2021, we aim at collaborating online. We want to bridge the LAKathon and LAK conference by explicitly inviting the LAK research sub-communities and Special Interest Groups to join and propose their challenges. The LAKathon 2021 intends to become the space for hands-on technical challenges, which take place in parallel to the work of the LAK sub-communities offering a space to address today's and tomorrow's challenges. For the first time in LAKathon history, there will be no space and physical presence constraints, thus offering new collaboration opportunities.

2 BACKGROUND

LAKathon 2015 focused on the *Apereo Open Dashboard* (Apereo, 2018), with data sourced from an Experience API (xAPI) Learning Record Store (LRS). It illustrated how the concept of an Open LA architecture can be made and discussed what a learning analytics dashboard must contain.

LAKathon 2016 explored *Open Learning Analytics*. Using as a reference point the emerging LA architecture developed by Jisc (Sclater, Berg, & Webb, 2015). The hackathon had a lasting effect, with numerous improvements to Jisc's interoperability recipes, setting the basis for greater integration of two emerging LA standards: Experience API and IMS Caliper (Edinburgh statement, 2016).

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 LA" (Duval, et al. 2015)) and "Data Literacy for LA". The second involved research on actionable analytics, student feedback, and embedding LA in pedagogic practice (Kitto, et al. 2016). The third involved the introduction of Jisc's student app, which was piloted with students across the UK.

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 (Di Mitri, 2018), and the creation of a Data Literacy Playground. The LAKathon 2018 also looked into algorithmic transparency and ethical workflows.

LAKathon 2019 revolved around three main challenges: the Interoperability Challenge which 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 the detection of disengagement. The third challenge which envisioned a markup language to describe blended learning courses was curriculum analytics.

LAKathon 2020, celebrated the 10th anniversary of the LAK conference, choosing the theme *"Accelerating Development by Learning from the Past"*. The key thematic objectives built upon the key topics of previous LAKathon editions and envisaged how Learning Analytics will generate impact on a ten-year timescale.

3 THEMES

The expected outcomes of the LAKathon are the identification and concrete pilot implementations of prototypes/tools/studies, which arise from the synthesis of educational technology, software development, and data science perspectives. As for previous events, the hackathon will generate repositories of code, sample data, screenshots, and slides from the activity of participants. At the LAKathon 2021, we expect to emphasise the following themes.

(A) Hacking the Hackathon: The organisers seek to strengthen, harden, and persist a virtual gathering of like-minded researchers, answering the questions: (1) Which novel practices associated with LA interventions do we instance? (2) How do we increase the positive influence of the LAK hackathons and better embed into the broader context of the discussion between the research and practitioner communities? (3) How do we accelerate the trajectory of research impacting the features and practices around Educational software?

(B) Labour Market-Education Divide. Current developments in the labour market pose a number of challenges on pockets of education, which aim at providing services to learners targeting particular skills needed for their careers. These educational settings are usually dynamic, personal, and specific to the type of work or skill, which is being targeted. To support this scenario, LA needs to go beyond traditional education-related processes and data sources. In this context sustainable career progression should be considered as a point of departure, and based on individual work-related objectives, personalized learning pathway(s) (curriculum(s)) should be built (Tavakoli et al, 2020).

(C) Video Conferencing Analytics. The newly enforced remote learning situation has made video conferencing the main channel of instruction in higher education. Most educators are simply replicating their traditional face to face classes into a video conferencing format. However, this format overlooks important aspects of education such as students' engagement. This challenge will explore the possibilities of conducting analytics on video conferencing data to explore good design practices to use video conferencing, potential applications that can support the instructor, and additional infrastructure or data-driven features that can improve the video conferencing experience of students (e.g. Seng & Ang, 2017).

(D) Psychological Ramifications of Learning in Isolation. Although the COVID crisis creates novel opportunities for LA, because more data are generated due to education taking place online, there is no evidence about the psychological ramifications of online learning. Isolation can bring about feelings of loneliness, disconnectedness, anxiety, and purposelessness, each of which may have a nontrivial impact on the individual learning process. By incorporating psychological measures of such constructs in the LA toolbox, we can understand which issues may need to be addressed to leverage the power of online learning by facilitating students' mental health.

(E) Digital Infrastructures. With the spread of the Covid19 pandemic, the majority of educational institutions have had to move their courses online. In most institutions "emergency teams" have been established to cope with the unprecedented mission of moving most of the education online. Online learning, however, does not equal emergency distance teaching. Robust digital infrastructures must be set in place when online teaching is blended with a physical presence. Video conferencing tools have become a widespread practice for online communication and collaboration. However, many small and medium educational institutions still lack adequate digital infrastructures to support their online learning initiatives including HW, LMSs, safe cloud storage, intranet channels for internal communication, netiquettes and privacy-preserving policies, etc.

(F) Diminishing the negative effects of Covid19, through the appropriate application of Learning paths. Vast seas of *OER material* exist of various qualities. We aim to support *different types of* isolated learners who are potentially suffering wider Covid isolation issues. This support is achievable by the group's definition, exploring gaps or underuse in current practices in Education and the Job market. Storytelling via the creation of personas. Later combining the personas with a systematic set of Covid support related learning paths. From these actions, we hope to focus effort through describable journeys that lead to the appropriate reuse of materials.

4 ORGANISATIONAL DETAILS

The LAKathon is organised as an online multi-day event. To overcome the time zone differences, the event will run asynchronously with a collaborative platform which will be open on the LAKathon website (www.lakathon.org). Each proposed theme will have a forum thread. The organisers will encourage subscribing to the platform and engaging in preliminary conversations before the event. In addition, through a Call for Proposals, we encourage the LA community to propose additional themes with short submissions (2 pages) detailing research questions, associated datasets, linked to the LAKathon themes. During the two days prior to LAK, there will be 3 synchronous online meetings (check-ins) at 3 different times to suit participants from around the globe. For the logistics, we need video conferencing facilities with a number of break out rooms (50 participants). The link-sharing and content management will be done via the platform and Git repository. The progress and outcomes will be disseminated via the website blog 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/18VJ9hVcfk9sOzs2MEeZ59Q5RSz1jZnUHzSWFSednEXg/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

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.

Tavakoli, M., Mol, S., & Kismihók, G. (2020). Labour Market Information Driven, Personalized, OER Recommendation System for Lifelong Learners. Proceedings of the 12th International Conference on Computer Supported Education, 96–104. https://www.scitepress.org/PublicationsDetail.aspx?ID=l5Pcp0p15O8=&t=1

Seng, K. P., & Ang, L. M. (2017). Video analytics for customer emotion and satisfaction at contact centers. IEEE Transactions on Human-Machine Systems, 48(3), 266-278.

LAK Theory 2021: 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 Gašević
Monash University
dragan.gasevic@monash.edu

ABSTRACT: The workshop addresses the enduring imperative of connecting theory and learning analytics – the ongoing work of striving for conceptual clarity, and being mindful of 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 (LA). This will be followed by a short plenary talk on the theory that works with big data. Participants will be invited to nominate a current research project or new research idea 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 and a template for an ongoing workshop initiative. To support the community, an online space will be created for ongoing collaboration.

Keywords: Research design, conceptual model, history of learning analytics, educational data

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 to 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 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. To do this involves addressing issues regarding definitions of concepts, design, model validation and interpretation of findings. Multidisciplinary groups of researchers working in the area need to come together to support this work and begin to create some common understandings in the field. This is the work proposed for the LAK 2021 theory workshop.

2 ORGANISATIONAL DETAILS

2.1 Half-Day Workshop Schedule

Table 1: Schedule.

Timing	Description	Contributors
5 minutes	Welcome and plan for today	Kathryn
20 minutes	'Setting the scene: Why focus on theory in learning analytics' 10 minutes presentation, 10 Q&A	Dragan
25 minutes	'Theory that works with big data' 15 minutes presentation, 10 Q&A	Sarah
40 minutes	Roundtable (Part 1). Work in progress roundtables: 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 [x2 before break]	Participants: 4 research teams per roundtable group
30 minutes	*BYO tea/coffee to a break out room*	*All*
40 minutes	Roundtable (Part 2). Continued [x2 after break]	Participants continued
30 minutes	Roundtable report back: group representatives to summarise conversation and potential impact on the work	Participants

20 minutes	Next steps plenary discussion, and close: (1) Further activities around theory and learning analytics: LAK 2022 workshop & book launch, LASI 2022 workshop/tutorial. (2) Interest in a mid-year check in?	Kathryn

2.2 Other Details

The event will be an open workshop. All attendees will have the opportunity to give a short presentation in their roundtable group on either work in progress or idea in development, should they wish to. Abstract submissions of 300-600 words for these short presentations will be handled via the event's Google Form: https://forms.gle/sbtATpZTjR5WzGCK6. 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: https://sites.google.com/view/lak21-theoryworkshop/home?authuser=0. 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.

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.

LAK21 Assess: Workshop on Learning Analytics and Assessment

Dragan Gašević and Mladen Raković
Monash University
{dragan.gasevic, mladen.rakovic}@monash.edu

Naif Aljohani
King Abudlaziz University
nraljohani@kau.edu.sa

José A. Ruipérez Valiente
University of Murcia
jruiperez@um.es

Sandra Milligan
University of Melbourne
s.milligan@unimelb.edu.au

Saeed Ul Hassan
Information Technology University
saeed-ul-hassan@itu.edu.pk

ABSTRACT: The workshop focuses on the connections between learning analytics and assessment. The intent of the workshop is to address some of the key open challenges in learning analytics that are related to reliability and validity of data collection and analysis, use of learning analytics in formative and summative assessment, measurement of learning progression, and assurance of assessment trustworthiness. The organizers will start the workshop by outlining links between learning analytics and assessment. An open call for contributions will be distributed to solicit brief descriptions of current research and practice projects for roundtable-style discussions with workshop participants. Expected outcomes are the formation of a community of practice and a possible follow-up publication.

Keywords: assessment, learning analytics, educational measurement

1 BACKGROUND

The field of learning analytics aims to harness the potential of digital traces of user interaction with technology. Through the analysis of digital traces, learning analytics seeks to advance understanding and support learning process, and improve environments in which learning occurs. Many promising results in learning analytics have promoted vibrant research and development activities and attracted much attention of policy and decision makers. To date, learning analytics demonstrated very promising results in several areas such as prediction and description of learning outcomes and processes (e.g., Baker et al., 2015; Gardner & Brooks, 2018; Greene et al., 2019), analysis of learning strategies and 21st century skills (e.g., Jovanović et al., 2017; Matcha et al., 2019), adaptive learner support and personalized feedback at scale (e.g., McNamara et al., 2012; Molenaar et al., 2012), and frameworks for ethics, privacy protection, and adoption (e.g., Tsai et al., 2018).

1.1 Challenge

Regardless of many promising results, the field still needs to address some critical challenges, including those at the intersection between learning analytics and assessment. For example, how can learning analytics be used to monitor learning progress? How can learning analytics inform formative and summative assessment as learning unfolds? In which ways can validity and reliability of data collection and analysis in learning analytics be improved? These challenges are of high significance in contemporary society that more and more requires development and use of complex skill sets (Greiff et al., 2017). Therefore, learning and assessment experience are closely associated. A growing body of research in educational data mining has been done on developing techniques that can support intelligent tutoring systems with the mechanisms for skill development (Corbett & Anderson, 1994; Desmarais & Baker, 2012). Yet, there is limited research that looks at how data collected and methods applied in learning analytics can be used and possibly constitute a formative or summative assessment. Moreover, can such data and methods satisfy requirements for assessments articulated in psychometric properties, methodological models, and different types of validity and reliability.

The role of learning analytics in analysis of assessment trustworthiness is another open research challenge. This has particularly been emphasized during the COVID19 pandemic with the emergency transition to distance and online education that also required different approaches to assessment that go beyond proctored exams. Several studies proposed the use of data analytic methods for detection of potential academic dishonesty and cheating behaviors. Although some interesting insights are ported and a strong potential to detect suspicious behaviors is demonstrated, there are many open challenges related to technical, ethical, privacy, practical, and policy issues of the development, implementation, and use of such data analytic methods.

1.2 Objective

The objective of this workshop is to promote research and practice that looks at the intersection of learning analytics and assessment. This workshop will examine approaches that build upon established principles in assessment to improve reliability, validity, usefulness of data collection and analysis in learning analytics. The workshop will also look into the ways how learning analytics can contribute to the future developments in assessment for both summative and formative purposes. The workshop will also examine practices for the use of learning analytics to assure assessment trustworthiness with the particular attention to the socio-technical nature of potential challenges.

2 ORGANISATIONAL DETAILS

2.1 Proposed Half-Day Workshop Schedule

Table 1: Proposed schedule.

Timing	Description	Contributors
5 minutes	Welcome and plan for today	Naif Aljohani

5 minutes	Links between learning analytics and assessment 5 minutes presentation, no Q&A	Dragan Gašević Mladen Raković
15 minutes	Using principles of assessment to improve quality of learning analytics 10 minutes presentation, 5 minutes Q&A	Sandra Milligan
15 minutes	Learning analytics and assessment trustworthiness 10 minutes presentation, 5 minutes Q&A	José A. Ruipérez Valiente
40 minutes	Roundtable (Part 1). Work in progress roundtables[1]: 10 minutes to introduce project, summarise progress to date, outline challenges to be overcome, and gather input from the group, followed by 10 minutes discussion with colleagues [x2 before break]	Participants: Four presentations
20 minutes	*Break and online socialization*	
40 minutes	Roundtable (Part 2). Continued [x2 after break]	Participants continued
20 minutes	Roundtable report back: group representatives to summarize conversation and potential impact on the work	Participants
20 minutes	Next steps plenary discussion, and close: Gauge interest in further activities around theory and learning analytics e.g. LAK 2022 workshop, LASI 2022 workshop/tutorial, mid-year check in, etc	Dragan Gašević

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 workshop's website. The submission timeline will follow the timeline suggested by the conference organizers, that is, call for participation 1 December 2020, deadline for abstract submissions 9 February 2021, and notification of acceptance 23 February 2021. We anticipate a registration of up to 30 participants. To reference this event on social media, the *#LAKAssess* should be used.

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. A possible follow-up publication will be organized in the form of a journal special issue.

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

[1] 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.

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, Background literature, Workshop materials, Working areas: Design, Method, Interpretation. Over time, as work develops and builds, additional resources will be provided to support ongoing development.

REFERENCES

Baker, R. S., Lindrum, D., Lindrum, M. J., & Perkowski, D. (2015). Analyzing Early At-Risk Factors in Higher Education E-Learning Courses. International Educational Data Mining Society (pp. 150-155)

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.

Desmarais, M. C., & Baker, R. S. (2012). A Review of Recent Advances in Learner and Skill Modeling in Intelligent Learning Environments. *User Modeling and User-Adapted Interaction*, *22*(1–2), 9–38.

Gardner, J., & Brooks, C. (2018). Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction*, *28*(2), 127-203.

Greene, J. A., Plumley, R. D., Urban, C. J., Bernacki, M. L., Gates, K. M., Hogan, K. A., ... & Panter, A. T. (2019). Modeling temporal self-regulatory processing in a higher education biology course. *Learning and Instruction*, 101201.

Greiff, S., Gasevic, D., & von Davier, A. A. (2017). Using process data for assessment in Intelligent Tutoring Systems. A psychometrician's, cognitive psychologist's, and computer scientist's perspective. *Design Recommendations for Intelligent Tutoring Systems. 5*, 171–179.

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., Uzir, N. A. A., Jovanović, J., & Pardo, A. (2019). Analytics of learning strategies: Associations with academic performance and feedback. In Proceedings of the 9th International Conference on Learning Analytics & Knowledge (pp. 461-470).

McNamara, D. S., Raine, R., Roscoe, R., Crossley, S. A., Jackson, G. T., Dai, J., ... & Dempsey, K. (2012). The Writing-Pal: Natural language algorithms to support intelligent tutoring on writing strategies. In Applied natural language processing: Identification, investigation and resolution (pp. 298-311). IGI Global.

Molenaar, I., Roda, C., van Boxtel, C., & Sleegers, P. (2012). Dynamic scaffolding of socially regulated learning in a computer-based learning environment. *Computers & Education*, *59*(2), 515-523.

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.

A Tutorial on Data Storytelling Techniques for Learning Analytics Dashboards

Vanessa Echeverria
Escuela Superior Politécnica del Litoral, ESPOL, Guayaquil, Ecuador
vecheverria@cti.espol.edu.ec

Lu Lawrence
Carnegie Mellon University, CMU, Pittsburgh, United States
llawrenc@andrew.cmu.edu

Yi-Shan Tsai, Shaveen Singh, Roberto Martinez-Maldonado
Monash University, Melbourne, Australia
{Yi-Shan.Tsai, Shaveen.Singh1, Roberto.MartinezMaldonado}@monash.edu

Gloria Fernandez-Nieto
University of Technology Sydney, Sydney, Australia
Gloria.M.FernandezNieto@student.uts.edu.au

ABSTRACT: Supporting educational stakeholders to interpret dashboards and visualizations poses critical design challenges that may often be trivialized. Teachers' and students' interpretation of visualized data is essentially the construction of a narrative about the learning process. Applying data storytelling techniques to design these visualizations can support the generation of insights derived from educational data by aligning the intended learning design, goals and outcomes with visual elements. The aim of this tutorial is to introduce participants to embrace data storytelling techniques into the design of visualizations and dashboards that can communicate meaningful insights.

Keywords: educational data storytelling, explainable dashboards, visual learning analytics

1 WORKSHOP BACKGROUND

1.1 Motivation

Although *learning dashboards* and other visual learning analytics (LA) have received significant traction in recent years (Bodily & Verbert, 2017; Schwendimann et al., 2017), there have also been numerous reports pointing to the limitations and possible pitfalls of rolling out these products without further research and development work (e.g. Jivet, Scheffel, Drachsler, & Specht, 2017; Teasley, 2017). Some of these limitations points to the absence of design choice justifications (Bodily & Verbert, 2017), poor evidence of grounding on educational theory (Jivet, Scheffel, Specht, & Drachsler, 2017), and the disaligment between teachers/students' needs and the learning analytics interfaces.

In parallel to these limitations, researchers and designers can easily overlook the learning context and the audience for whom these visualisations have been created (Schwendimann, et al. 2017). Sometimes, designers and researchers want to communicate multiple insights or dimensions of data about students' experience. The conventional approaches adopted by researchers and designers can lead to the design of overly complex visualisations that are often hard to interpret (Duval 2011) especially "at a glance". Moreover, teachers and students are commonly encouraged to interpret these visualisations in a limited time due to other activities happening at the same time and, even if the data can be interpreted correctly, they may fail to understand how to act upon such data, failing to adapt their behaviour (Greller & Drachsler 2012). A major challenge for learning analytics researchers and developers is to support the discovery and communication of insights, for students and teachers not needing to play the role of data analysts, at the risk of gaining no insight.

This tutorial focuses on data storytelling, the ability to convey data not just in numbers or charts, but as a narrative that the audience can comprehend using storytelling foundations (e.g. plots, twists and calls to action; Lee, Riche, Isenberg and Carpendale (2015)). Data storytelling (DS), which builds on classic InfoVis guidelines (Tufte & Schmieg 1985), is a structured approach for communicating data insights, and it involves a combination of three key elements: *data, visuals, and narrative* (Dykes 2015). Narrative helps explain what the data and visualizations are conveying and why particular insights are important. Prior work on LA community has explored how these DS elements play an important role in supporting the understanding of complex learning data and how these elements drive teacher's attention when aligned to expected outcomes or learning goals (Echeverria et al. 2018a, 2018b). In short, this tutorial brings a practical approach for applying data storytelling principles to address the analytical challenge of visualizing complex and heterogeneous data and facilitating the communication of insights.

1.2 Objectives

One of the key goals of this interactive tutorial is to bring researchers, practitioners, and other educational stakeholders into a design space to provide a set of tools/methods for handcrafting visualizations that are relevant to the context by guiding the user's attention to key insights (i.e. derived from the learning design/expected outcomes). This tutorial will enable researchers and practitioners to apply data storytelling techniques into their practice when designing learning dashboards.

Main activities of this tutorial will include: (1) an introduction to data storytelling tools and methods, (2) a hands-on activity for designing of a lo-fi prototype of the participant's visualization or dashboard that includes storytelling elements, and (3) networking opportunities with researchers in the field. Finally, it is also expected to build a community, particularly for educational data storytelling research.

2 PROGRAM

2.1 Schedule

The following activities have been planned for a half-day tutorial:

1. Introductions (30 mins)
2. Presentation: data storytelling (40 mins)
3. Discussion (20 mins)
4. Break - social gathering (15 mins)
5. Guided activity (Part 1): working with data visualizations and data storytelling (40 mins)
    a. Identifying potential stories/insights
    b. Choosing the visualization that fits your data
    c. Linking visual elements with stories/insights
6. Lunch - networking (40 mins)
7. Guided activity (Part 2): working with data visualizations and data storytelling (40 mins)
    a. Linking visual elements with stories/insights
    b. Presentations
8. Concluding remarks (15 mins)

2.2 Participants and recruitment

This is an open tutorial. Participants will be required to register to attend the workshop. We expect at least 20 participants to attend the workshop. We plan to recruit participants through social media (i.e. Facebook, twitter). In addition, we will create a webpage with all relevant information about the tutorial.

2.3 Materials and equipment

Now that the conference has moved to an online format, we will provide prior material in advance for participants. This will help participants to revise the material in advance if they are not able to participate in any part of the tutorial due to time differences.

Zoom or any other video conferencing system will be used as a means of communication.

Authors of this tutorial will act as mentors and will provide assistance during hands-on activities.

We will use different online tools to work collaboratively (i.e. miro, google slides) and additional tools to provide social presence and networking (i.e. gather.town).

Two types of participation are expected for the guided activity: 1) participants will bring some sort of visualization or dashboard (it can be a low-fidelity prototype) to work on; or 2) we will provide some examples they can work on.

REFERENCES

Bodily, R., Kay, J., Aleven, V., Jivet, I., Davis, D., Xhakaj, F., & Verbert, K. (2018). Open learner models and learning analytics dashboards: a systematic review. In Proceedings of the *Proceedings of* *the 8th International Conference on Learning Analytics and Knowledge*, (pp. 41-50). Sydney, New South Wales, Australia. 3170409: ACM.

Duval, E. (2011). Attention please!: learning analytics for visualization and recommendation. In Proceedings of the *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*, (pp. 9-17). Banff, Alberta, Canada. New York: ACM.

Dykes, B. (2015). Data storytelling: What it is and how it can be used to effectively communicate analysis results. *Applied Marketing Analytics,* 1(4), 299-313.

Echeverria, V., Martinez-Maldonado, R., Granda, R., Chiluiza, K., Conati, C., & Buckingham Shum, S. (2018a). Driving Data Storytelling from Learning Design. In Proceedings of the *Eighth International Learning Analytics & Knowledge Conference*, (pp. 131-140). ACM.

Echeverria, V., Martinez-Maldonado, R., Buckingham Shum, S., Chiluiza, K., Granda, R., & Conati, C. (2018b). Exploratory versus Explanatory Visual Learning Analytics: Driving Teachers' Attention through Educational Data Storytelling. *Journal of Learning Analytics*, *5*(3), 72—97. https://doi.org/10.18608/jla.2018.53.6

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology & Society,* 15(3), 42–57.

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). *Awareness is not enough: pitfalls of learning analytics dashboards in the educational practice.* Paper presented at the European Conference on Technology Enhanced Learning, (pp. 82-96). Springer.

Lee, B., Riche, N. H., Isenberg, P., & Carpendale, S. (2015). More Than Telling a Story: Transforming Data into Visually Shared Stories. *IEEE Computer Graphics and Applications,* 35(5), 84-90.

Reimers, G., & Neovesky, A. (2015). Student focused dashboards. In Proceedings of the *7th International Conference on Computer Supported Education (Lisbon)*, Lisbon.

Schwendimann, B. A., Rodríguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., & 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.

Teasley, S. D. (2017). Student Facing Dashboards: One Size Fits All? *Technology, Knowledge and Learning,* 22(3), 377-384.

Tufte, E. R., & Schmieg, G. M. (1985). *The visual display of quantitative information* (Vol. 53): American Association of Physics Teachers.

LAK21 Playful Collaboration: Workshop on Design of Learning Analytics for Digital Game Use in the Classroom

Yoon Jeon Kim[a], José A. Ruipérez-Valiente[b], Grace C. Lin[c], Nathan Holbert[d], Matthew Berland[a], Baltasar Fernández Manjón[e], and David Gagnon[a]
[a] University of Wisconsin-Madison, [b] University of Murcia, [c] Massachusetts Institute of Technology, [d] Teachers College, Columbia University, [e] Complutense University of Madrid
{yj.kim, mberland, djgagnon}@wisc.edu, jruiperez@um.es, gcl@mit.edu, holbert@tc.columbia.edu, balta@fdi.ucm.es

ABSTRACT: The workshop examines how we can leverage the existing theories in game-based learning and teacher assessment literacy to develop learning analytics and data visualizations for game-based learning. In particular, it focuses on the affordances and challenges of collaborating with educators to make the games and data useful and meaningful for classroom use. The organizers will introduce a process that can lead collaborative development of learning analytics and kickoff the workshop with an overview presentation of their four games. Participants of the workshop will get to deep dive with one game of their choosing. The remainder of the workshop will showcase participant presentations selected from an open call for contributions. Participants will get to present, discuss, and receive feedback from one another. We anticipate a Special Interest Group emerging from the workshop as well as potential publications.

Keywords: games for learning, learning analytics, codesign

1 BACKGROUND

The educational benefits of games have been well documented over the past decade. In a recent meta-analysis, Clark and his colleagues (Clark et al., 2016) report that compared to nongame conditions, games had a moderate to strong effect for improving overall learning outcomes including cognitive and interpersonal skills. Another review (Boyle et al., 2016) similarly reports games as beneficial for learning across domains such as knowledge acquisition, affect, behavior change, perception and cognition as well as 21st Century Skills. While ample evidence shows that games, in general, have a great potential to support learning, only when combined with a thoughtful curriculum considering teacher's practices and classroom-contexts, can they be successfully used to support learning in classrooms (Klopfer et al., 2009). Additionally, the vast and rapid data generation from learners' interactions with the game can be overwhelming for the educators to manually process, thus possible new insights about the learners can remain unavailable. Learning analytics holds the potential to transform the game data into meaningful information that can be used by the teacher as part of the formative assessment process and to provide students with detailed and individualized support and feedback. Therefore, learning analytics and appropriate visualizations coupled with learning games can provide opportunities for awareness, reflection, and sense-making, thereby improving learning and increasing the impact of games in classrooms.

1.1 Challenge

However, it is a challenge to implement learning analytics (LA) in digital games such that they are grounded in theory and practice, technically sound, and useful for teachers without flashy, cumbersome additional features. There are multiple angles of this challenge for the LA community to address. First, we need to understand teachers' assessment literacy in the context of LA in games. Teachers are not necessarily fluent with the technical concepts to critically evaluate learning analytics, and they may also have difficulty in understanding how the models are developed, validated, and implemented. The "black-box" nature of learning analytics, in which teachers are asked to simply trust the outputs without understanding the intricacies may lead to mistrust, uncertainty, or misuse of the models and their results (Rudin, 2014). Though work has begun to improve the interpretability of LA (Calvo-Morata *et al*., 2018), less attention has been paid to investigating teachers' assessment literacy. Second, we need to consider teachers' pedagogical approaches and goals as these can drive their data and information needs from the game. The uncertainty surrounding LA's pedagogical relevance and the level of engagement with pedagogies have been well documented (see Tan & Koh, 2017), but LA community cannot remain pedagogically agnostic; LA models need to be designed with teachers' pedagogical purposes and goals in mind. Finally, in line with pedagogical considerations, broader ecological validity of LA in the classroom should be addressed. Games offer the unique affordances to peek at nuanced constructs beyond proficiency or task completion, and teachers interested in integrating games in their curriculum are likely to value these outcomes.

1.2 Objective

To address these challenges and meet the goal of having a real impact in classrooms, researchers need to develop LA that are meaningfully connected with learning theories as well as reflect teachers' classroom practices. We propose an interactive workshop that bring diverse methods and approaches to productively collaborate with teachers for the meaningful selection and development of learning analytics as well as data visualizations in digital game environments. This workshop brings together four games for learning projects (see Table 1) for the audience (a) to explore different affordances of each game environment regarding the stories one can tell about the learner from the game telemetry data (i.e. what to measure?), (b) to discuss with the teams the classroom contexts that they have envisioned (i.e. for what pedagogical and instructional decisions), and, therefore, (c) how they approached creating the learning environment as well as learning analytics, and (d) how the teachers were involved as collaborative partners in the process.

Table 1: Educational Game Descriptions.

Game	Description	Contributors
Shadowspect	A 3D geometry game for assessing students' spatial reasoning. Uses iterative cycles to codesign with teachers to build pertinent metrics (e.g., on student persistence) and data visualizations	3 organizers
Beats Empire	A music management game where players build their own music empire by interpreting data on listener interests to make decisions about what artists to sign and what songs to record. It includes a dashboard system that provides teachers with rich data about how players use data representations in the game and suggested actions for encouraging players to go deeper with target data analysis concepts.	2 organizers
Conectado	A video game to raise awareness on bullying and cyberbullying among 12- to 17-year-olds through experience and emotions. It is intended to help educators increase the interactivity and emotional engagement of their classes.	1 organizer

Lakeland	A real-time strategy game that teaches the nutrient cycle between humans, corn farms, dairy farms and lake ecosystems. This game also includes a teacher-facing dashboard of descriptive and predictive analytics to support game facilitation.	1 organizer

We will ground the work by introducing a process (see Figure 1) that can lead a collaborative development of learning analytics by considering various contextual factors—e.g., assessment or pedagogical practices that researchers and developers intend to support, data and assessment literacy of the target audience, data affordances of the game environment.

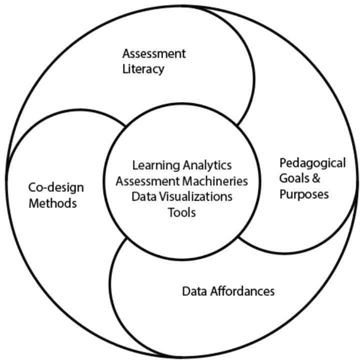


Figure 1: Conceptual Framework for the Collaborative Development Process

Using the mechanics detailed in Section 2, the audience will become familiarized with a process that they can apply to collaboratively design and develop learning analytics with classroom teachers.

2 ORGANISATIONAL DETAILS

2.1 Proposed Half-Day Workshop Schedule

Table 2: Proposed schedule.

Timing	Description	Contributors
15 minutes	Framing and Introduction	Organizer1
40 minutes	Whole group overview: 10 minute mini presentations from each of the 4 game groups	Organizers
30 minutes	Deep Dive Breakout Rooms: Discussions and Q&A	Everyone
15 minutes	Stretch break	
40 minutes	Roundtable Breakout Rooms (Part 1): There will be 4 groups in each room. Each group will get 10 minutes to introduce project and the challenges they are facing (or anticipating) and 10 minutes for feedback and discussion with organizers and other participants. [x2 before break]	Participants
15 minutes	Break and online socialization	
40 minutes	Roundtable Breakout Rooms (Part 2). Continued [x2 after break]	Participants
20 minutes	Roundtable Share-Out: Each breakout room's representatives will share insights from the conversation with the participants	Participants
15 minutes	Closing remarks: Organizers will outline the next steps, soliciting feedback and gauge interest from participants in forming a SIG and a special issue on the themes around collaborating with teachers on games in classrooms as well as a positioning paper	Organizer1

2.2 Other details

The workshop will be an open event. All attendees will be able to participate in discussions, but only selected projects will present during the roundtables. Interested groups or individuals can submit abstracts for the roundtable presentations. The submission deadline will follow that suggested by LAK conference organizers. We anticipate registration of up to 25 participants. Please use #LAK_teachersgames when referencing this event on social media.

3 OBJECTIVES/INTENDED OUTCOMES

By attending this workshop, the participants will gain a deeper understanding and sense of the practical implications of the learning analytics of games for classroom purposes. They may become aware of potential challenges in this work, and we hope that they will take away a set of conceptual tools and theoretical frameworks and approaches that may guide their future work in the context of playful collaborative design of digital games in the classroom. We will document the outcomes of the workshop on the workshop website, which will also serve as the starting point of connection of a potential Special Interest Group in this area as well as special issue publications.

4 WEBSITE STRUCTURE AND CONTENT

The team will create a Google website that serves the following purposes. Before the workshop, potential contributors can read more about the intentions and goals for this workshop and four existing projects to gain background knowledge about how each team has been addressing this issue in their projects. Materials to be provided include links to the games, conceptual frameworks and decision-making tools, and examples of co-design activities. During the workshop, the website will function as a central space to document and gather artifacts and group interactions. After the workshop, the website will be used to disseminate the insights gained from the workshop as well as a place to promote community building around the themes that emerged from the workshop.

REFERENCES

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., Lim, T., Ninaus, M., Ribeiro, C., & Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education, 94*, 178–192. https://doi.org/10.1016/j.compedu.2015.11.003

Calvo-Morata, A., Alonso-Fernández, C., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2018). Making Understandable Game Learning Analytics for Teachers. In *International Conference on Web-Based Learning* (pp. 112–121). Springer.

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research, 86*(1), 79–122.

Klopfer, E., Osterweil, S., & Salen, K. (2009). *Moving learning games forward.* The Education Arcade.

Rudin, C. (2014). Algorithms for interpretable machine learning. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 1519–1519). ACM.

Tan, J. P.-L., & Koh, E. (2017). Situating learning analytics pedagogically: Towards an ecological lens. *Learning: Research and Practice, 3*(1), 1–11. https://doi.org/fgnm

LALN: Building Capacity for Learning Analytics

Justin T. Dellinger
University of Texas at Arlington
jdelling@uta.edu

George Siemens
University of Texas at Arlington and University of South Australia
gsiemens@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

ABSTRACT: The Learning Analytics Learning Network (LALN) was created in 2019 as a scalable way to build capacity for learning analytics by leveraging existing communities of practice around the world. LALN events are held throughout the year and speakers from different local or regional learning analytics communities take turns to introduce participants to new methods and activities. This half-day workshop will start with a presentation by the organizers on the background of LALN. Next, a panel of leaders from different communities will discuss the challenges and opportunities of this network. Finally, participants will work in groups to explore how to best capitalize on this network and discuss their needs.

Keywords: Learning Analytics; International Network; Learning Networks; Data Science; Community of Practice; Networks of Practice

1 INTRODUCTION TO THE LEARNING ANALYTICS LEARNING NETWORK

Although data science has emerged as an important part of educational research and practice in order to meet the demand for a workforce that has literacy in data science methods and competence educational data, research, and practice (Baker & Siemens, 2014), there is an insufficient number of graduate programs and other professional development and training activities to satisfy this need. As a result, the larger learning analytics workforce lacks key competencies. To address this gap, a consortium of scholars from the University of Texas at Arlington, University of Pennsylvania, and University of South Australia have developed the Learning Analytics Learning Network (LALN).

Meetings are held worldwide in local and regional hubs that take turns hosting a distinguished speaker and streaming the event online so other cities can join synchronously. All events are also recorded for later viewing to accommodate the different time zones and are complemented by resources for facilitating sessions. Local moderated discussions are also held when in-person events are possible. The activities and exercises range from beginner to advanced and target different audiences. 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. The activities function as both 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

The organizers will begin the workshop with a presentation on the background of LALN. They will discuss the challenges of moving research into practice and developing the requisite skills to effective make use of learning analytics approaches. Finally, they will also address the vision, model, and operation of the network. The goal of this part of the workshop is to help set the context for subsequent group brainstorming and discussions.

3.2 Panel discussion

The organizers have reached out to leaders in local and regional learning analytics communities to participate in this part of the workshop. These leaders will discuss the opportunities and challenges of hosting LALN events. They will also discuss how to build communities and run sessions based on their experience.

3.3 Brainstorming session

Session attendees will then participate in a brainstorming session to explore and detail the needs that LALN could support. The larger group will discuss topics such as the strategies used by different organizations, the types of events and how often they should occur, and how best to facilitate networking opportunities to build connections and support the growth of the field. In this session, attendees will work in small groups (3 to 5 people) and will share out to the larger group at the conclusion.

4 WORKSHOP FORMAT

This half-day open workshop can accommodate up to 40 participants. The organizers plan to create a website to share any resources generated before and during the workshop (e.g., key readings, presentations, video clips, discussion notes, and documentation for joining the network). Given that LAK21 will take place fully online, we will make use of the web conferencing platform for whole- and small-group sessions. The organizers have significant experience with this format.

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.

Human Centred Learning Analytics

Roberto Martinez-Maldonado
Monash University
roberto.martinezmaldonado@monash.edu

Yannis Dimitriadis
Universidad de Valladolid
yannis@yllera.tel.uva.es

Kenneth Holstein
Carnegie Mellon University
kjholste@cs.cmu.edu

Alyssa Wise
New York University
alyssa.wise@nyu.edu

Carlos Prieto-Alvarez
The University of Sydney
carlos.prietoalvarez@sydney.edu.au

Fabio Campos
New York University
fcc261@nyu.edu

Juan Pablo Sarmiento
New York University
jps651@nyu.edu

June Ahn
University of California, Irvine
junea@uci.edu

Lu Lawrence
Carnegie Mellon University
llawrenc@andrew.cmu.edu

Simon Buckingham Shum
University of Technology Sydney
Simon.BuckinghamShum@uts.edu.au

ABSTRACT: The term human-centred learning analytics (HCLA) was recently coined to refer to the subcommunity of LA researchers and practitioners interested in utilising the body of knowledge and practice from design communities, such as participatory design and co-design, into data-intensive educational contexts. Although there is a growing interest in designing LA systems with students and teachers, several questions still remain regarding how the LA community can appropriate design approaches from other communities and identify best practices that can be more suitable for LA developments. This workshop intends to address some of these questions.

Keywords: design, human-centred, co-design, participation,

1 INTRODUCTION

This workshop seeks to build on the momentum from recent years within the LAK community, around the contributions that Human-Centred Design theory and practice should make to Learning Analytics system conception, design, implementation and evaluation. The theme of LAK18 was *Towards User-Centred Design*, where there were two sessions devoted to this topic (LAK18-UCD, 2018). At this conference, the first *LAK Participatory Design workshop* was convened, providing an identity to this particular strand of work (Prieto-Alvarez et al., 2018). This was consolidated in a special issue of the Journal of Learning Analytics on *Human-Centred Learning Analytics* (Buckingham Shum, et al. 2019), and four PhDs have recently been completed with explicit attention to participatory design for LA, reflecting interest in the emerging generation of researchers (Dollinger, 2019; Echeverria, 2020; Holstein, 2019; Prieto-

Alvarez, 2020). Moreover, the best practitioner paper at LAK20 focused on co-designing with learners (Sarmiento et al., 2020).

1.1 Background

(Mis)understandings of real-world users, stakeholders, contexts, and routines can make or break LA tools and systems. However, the extent to which existing human-centred design methods, processes, and tools are suited to address such human and societal factors in the context of LA is a topic that remains under-explored by our community. In response, the term human-centred learning analytics (HCLA) was recently coined (Buckingham Shum et al., 2019) to refer to the subcommunity of LA researchers and practitioners interested in utilising the body of knowledge and practice from design communities, such as participatory design and co-design, into data-intensive educational contexts. Holstein et al. (2017) were the first in adapting various co-design techniques to identify teachers' data needs and build prototypes of awareness tools with them. In fact, teachers have been the most commonly involved stakeholders in LA co-design studies. For example, Dollinger et al. (2019) discussed implications for the use of participatory semi-structured interviews with teachers in long-term LA projects. Wise and Jung (2019) combined LA interface walkthroughs and transcript analysis to make design decisions for a dashboard intended to be used by teachers. Prestigiacomo et al. (2020) explained how generative tools can be used to investigate the broad challenges that teachers are facing to then focus on those that can be addressed by automatically generating evidence for reflection. Holstein et al. (2019) featured a number of co-design techniques, namely card sorting, directed storytelling, semi-structured interviews, prototyping and behavioural mapping, to co-design a classroom analytics innovation with teachers. Whilst some examples of LA design processes have focused on engaging with students, these are just starting to emerge (Chen & Zhu, 2019; de Quincey et al., 2019; Prieto-Alvarez et al., 2018, Prieto et al., 2019; Sarmiento et al., 2020).

1.2 Aim of the Workshop

The studies presented above make it evident that there is a growing interest in designing LA systems with students and teachers. But several questions remain regarding how the LA community can appropriate design approaches from other communities and identify best practices that can be more suitable to LA developments. However, little work has been done in proposing the steps that other researchers or designers can use as a guidance to structure participatory sessions to understand critical aspects of the envisaged use of LA tools and the actual data needs that stakeholders may have. This workshop aims at consolidating the subcommunity of LA researchers and practitioners interested in the human factors related to the effective design of LA innovations. In doing so, we plan to address questions such as: What has been done so far in HCLA, and what have we learned from these experiences? Within the context of our field, how do we define some fuzzy concepts such as "participatory", "co-design" and "human-centeredness"? Finally, as a community, what do we want to know (research agenda) from now on?

Thus, the intended outcome of this workshop is twofold:

Outcome 1: A plan for the consolidation of a new SoLAR SIG dedicated to the study and practice of HCLA within the larger LAK community; and

Outcome 2: The publication of a report summarising the workshop experience and, hopefully, a "roadmap manifesto" setting a research agenda for HCLA.

2 ORGANISATIONAL DETAILS

2.1 Workshop format, participation, and pre-workshop task

The workshop is envisioned to be a half-day, fully online workshop. Between 12 and 24 participants, with a shared interest in human-centred learning analytics, are expected to be part of this workshop. This workshop welcomes everyone with an interest in the field, from beginners to experts. We will not have a call for papers. Instead, participants will be asked to fill a survey which will capture previous experiences in HCLA as well as current understandings of design aspects that will be relevant for the discussions during the workshop. In particular, participants will be asked to share their experience with human centred design or human centred LA; to define human-centred design; to share what design methods they are familiar with; future plans to adopt human-centred design methods in LA projects.

2.2 Workshop activities

The workshop is planned to take place during the pre-conference activities of the main conference and is planned for a half-day format of up to 4 hours (April 11 or 12, 2021). The workshop is divided into four parts:

Overview of HCLA. In the first part of the workshop, and based on the survey results, we will present a number of processes, frameworks and examples for engaging in participatory/co-design processes with students, faculty or administrators, emphasising both opportunities and challenges.

Human-Centred Design challenge. The second part of the workshop is a collaborative *design challenge*. Participants will engage in creating a research design plan by using human-centred methodologies. They will be grouped in teams of 4-5 people, and go to virtual breakout rooms. They will be presented with a design need and asked to work together designing a human-centred design project to handle the need. Groups will be prompted to consider methodologies such as Value-Sensitive Design, Co-design and generative tools in planning their projects.

Sharing and guided critique. The third part will be a discussion based on the experience co-designing the human-centred plans. A number of discussants from other communities (e.g. human-computer interaction, interaction design, participatory design and information visualisation), and some that critique human-centred design methods, will be invited to the workshop to give their critical points of view on the ideas posed in the design plans. We expect that this will lead to a discussion of the pros and cons of human-centred design techniques, what needs to be adapted to fit LA purposes and the differences of meaning of human-centred design for different people.

Discussion on next steps. All participants will be invited to contribute with ideas to set a potential HCLA research agenda and the potential configuration of a dedicated SIG.

2.3 Dissemination strategy

A workshop website will be made available upon acceptance of this workshop. A call for participation will be generated and published via the website, and through the twitter accounts and mailing lists the workshop organisers have access to. The website will also include an overview of the aims of the workshop, information about the workshop organisers, contact details and reports and other outputs from the workshop.

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.

Buckingham Shum, S., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-Centred Learning Analytics. *Journal of Learning Analytics,* 6(2), 1-9.

Dollinger, M., Liu, D., Arthars, N., & Lodge, J. (2019). Working Together in Learning Analytics Towards the Co-Creation of Value. *Journal of Learning Analytics, 6*(2), 10–26.

Echeverria, E. (2020). *Designing Feedback for Collocated Teams using Multimodal Learning Analytics.* Doctoral Dissertation, Connected Intelligence Centre, University of Technology Sydney, Australia.

Holstein, K., McLaren, B. M., & Aleven, V. (2017). Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms. In *Proceedings of the seventh international learning analytics & knowledge conference* (pp. 257-266).

Holstein, K. (2019). *Designing Real-time Teacher Augmentation to Combine Strengths of Human and AI Instruction.* Doctoral Dissertation, Human-Computer Interaction Institute, Carnegie Mellon University, USA.

Holstein, K., Hong, G., Tegene, M., McLaren, B. M., & Aleven, V. (2018). The classroom as a dashboard: co-designing wearable cognitive augmentation for K-12 teachers. In *Proceedings of the 8th international conference on learning Analytics and knowledge* (pp. 79-88).

LAK18-UCD (2018): User-Centred Design Sessions I and II, *Proceedings of the 8th International Conference on Learning Analytics.*

Prestigiacomo, R., Hadgraft, R., Hunter, J., Locker, L., Knight, S., van den Hoven, E., & Martinez-Maldonado, R. (2020). Learning-centred translucence: an approach to understand how teachers talk about classroom data. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 100-105).

Prieto, L. P., Rodríguez-Triana, M. J., Martínez-Maldonado, R., Dimitriadis, Y., & Gašević, D. (2019). Orchestrating learning analytics (OrLA): Supporting inter-stakeholder communication about adoption of learning analytics at the classroom level. *Australasian Journal of Educational Technology*, *35*(4).

Prieto-Alvarez, C. (2020), *Engaging Stakeholders in the Learning Analytics Design Process.* Doctoral Dissertation, Connected Intelligence Centre, University of Technology Sydney, AUS.

Prieto-Alvarez, C., Martinez-Maldonado, R., & Shum, S. B. (2018). *Mapping Learner-Data Journeys: Evolution of a Visual Co-Design Tool*. Paper presented at the ACM Australian Computer-Human Interaction Conference, OzCHI'18.

Sarmiento, J. P., Campos, F., & Wise, A. (2020). Engaging Students as Co-Designers of Learning Analytics. *Companion Proceedings 10th International Conference on Learning Analytics & Knowledge* (LAK20)

Wise, A. F., & Jung, Y. (2019). Teaching with Analytics: Towards a Situated Model of Instructional Decision-Making. *Journal of Learning Analytics, 6*(2), 53–69.

Accessible Learning, Accessible Analytics: a Virtual Evidence Café

Papathoma Tina, Rebecca Ferguson, Dimitrios Vogiatzis
The Open University, UK
Tina.Papathoma@open.ac.uk, Rebecca.Ferguson@open.ac.uk,
Dimitrios.Vogiatzis@open.ac.uk

ABSTRACT: Learner accessibility is often thought of in terms of physical infrastructure or, in the case of online learning, guidelines for web design. Learning analytics offer a new set of possibilities for identifying and removing barriers to accessibility in learning environments. This is not simply a matter of designing analytics tools to be more accessible, for example by catering for learners who need extra time to respond, reducing cognitive load, or choosing an appropriate colour palette. When it comes to increasing access to learning opportunities for people with disabilities, solutions must be developed in the field of learning analytics. This workshop is a step towards developing those solutions. It will take the form of an evidence café, a structured event in which participants will be split into groups to discuss technical and pedagogic approaches to accessibility, as well as the barriers faced by disabled students and educators, and the associated challenges faced by those who design and research learning analytics. The intended outcomes of this workshop are to raise awareness of accessible learning and accessible analytics, and to build a community of researchers to lead future development in the area of accessible analytics.

Keywords: Accessibility, disability, evidence café, inclusion, learning analytics

1 BACKGROUND

Students with disabilities are less likely than other students to complete their studies, go on to complete higher degrees, or secure graduate employment (Mamiseishvili & Koch, 2012). This disparity is evident when large datasets are used to examine success and completion rates, segmenting findings using demographic filters. In such studies, many of which predate the emergence of learning analytics as a field, disability is one variable among many (Tinto, 1997).

Learning analytics, with its goals of 'understanding and optimising learning and the environments in which it occurs' (Long & Siemens, 2011, p34) offers the possibility of identifying and removing barriers to accessibility in learning environments. There are two elements to this work. First, it is important that learning analytics tools do not introduce new accessibility issues. Second, analytics should be designed to increase access to learning opportunities.

Accessibility is often thought of in terms of the guidelines set out by the W3C web standards body, and that web accessibility means 'people with disabilities can perceive, understand, navigate, and interact with the Web, and that they can contribute to the Web' (W3C, 2018). By extension, learning analytics tools and dashboards can be considered accessible if people with disabilities can perceive, understand, navigate, interact with them and contribute to them. One aspect of this work is technical – designers will take accessibility into account when adding buttons and visual elements, deciding on colours, contrast, fonts and font size. Another aspect requires more thought about usability; does the tool cater for learners who need extra time to make responses; reduce cognitive load as far as possible; and remove triggers of anxiety? (Lister et al, 2020)

General principles and standards may be applied to technical and usability elements. When it comes to increasing access to learning opportunities for people with disabilities, solutions must be developed in the field of learning analytics. Three strands of work indicate potential ways forward. Work presented at LAK16 indicated some of the ways in which analytics might be used to contribute to disabled students' learning, initially by using large datasets to identify courses on which students with declared disabilities had significantly lower success rates than other students (Cooper, Ferguson, & Wolff, 2016).

Development of conversational user interfaces (chatbots) has highlighted how language – use of jargon, overuse of abbreviations, and the introduction of confusing terms – can all present barriers to learners with cognitive disabilities, mental health issues, or some types of learning difficulty (Lister, Coughlan, Iniesto, Freear, & Devine, 2020). Elsewhere, work on serious games has pointed to ways in which learning analytics could be used to provide support for people with intellectual disabilities, personalising learning pathways and flagging when key areas of content have not been accessed (Nguyen, Gardner, & Sheridan, 2018).

1.1 Motivation

This workshop will identify ways of making both analytics and learning more accessible to people with disabilities. Disability is here taken to relate to 'barriers created by catering to assumptions about what most people can do. Disabilities include physical, cognitive, motor or mental difficulties/impairments, as well as barriers associated with factors such as dyslexia and age. People also face barriers when a course is not in their preferred language. Disability may involve technological or pedagogical barriers to learning' (Papathoma, Ferguson, Iniesto, Rets, Vogiatzis & Murphy, 2020).

1.2 Relevance to the conference theme

This workshop relates directly to the contribution that learning analytics can make to learning. It is concerned with an ethical aspect of learning analytics, the need for an equitable approach that takes account of the needs of all learners.

2 WORKSHOP OBJECTIVES AND INTENDED OUTCOMES

The workshop objectives are to explore the elements about both technical and pedagogic approaches to accessibility, as well as the barriers faced by disabled students and educators, and the associated challenges faced by those who design and research learning analytics. The Evidence Café approach seeks to bring together researchers and practitioners, those with theoretical knowledge and those with expert practical knowledge; those who have encountered barriers to accessibility and those who are working to remove those barriers. The intended outcome is to raise awareness to researchers and practitioners to accessible learning and accessible analytics with an underlying goal to build a community of researchers to lead future development in the area of accessible analytics.

3 WORKSHOP ORGANISATION

This will be a half-day open interactive workshop event, open to any interested LAK delegate, which will take the form of an Evidence Café. These are informal workshop-style events where expert participants are split into groups to discuss an issue guided by a discussion object, which is used to facilitate meaningful conversations between practitioners and academics (Clough, Adams, & Halford, 2017). The discussion object gives participants a shared language to discuss the topic at hand. In order to complete the associated facilitated activities each participant must have the opportunity to voice their thoughts. This participatory method supports the translation of research into practice, supporting a deep understanding of the use of evidence in practice, and providing a forum for knowledge exchange. The workshop seeks to have 3 groups of 5 attendees maximum so that it will be possible to facilitate online discussions.

The Evidence Café approach is well suited to LAK because it is a workshop approach that is designed to bring together researchers and practitioners, those with theoretical knowledge and those with expert practical knowledge; those who have encountered barriers to accessibility and those who are working to remove those barriers. It provides opportunities for structured conversations and the sharing of knowledge, and its combination of informal interaction and facilitated discussion has been shown to work well in an online environment (Papathoma et al., 2020).

In this workshop, the discussion object will be an accessibility analysis of an existing learning analytics tool. This will be used to prompt discussions about both technical and pedagogic approaches to accessibility, as well as the barriers faced by disabled students and educators, and the associated challenges faced by those who design and research learning analytics.

3.1 Schedule

In this half day workshop/evidence café the proposed schedule includes

- a short introduction to evidence cafés (30 min)
- an introduction to each other -ice breaker activity (30 min)
- Activity 1: What are the technical and pedagogic approaches to accessibility in the Learning Analytics tool example you were given as a discussion object? (30 min discussion amongst individual groups and 20 min sharing with all groups)
- Activity 2: What enables and/or hinders the accessibility analysis of the learning analytics tool you were given as a discussion object? (30 min discussion amongst individual groups and 20 min sharing with all groups)
- Wrap up (10 min)
- Feedback on the evidence café (10 min)

3.2 Organisers

A group of scholars with previous experience of organizing workshops and Evidence Cafés in virtual settings. We come from institutions that are committed to making learning more accessible.

3.3 Workshop website and publicity

The workshop will be publicized using social media, including both Twitter and LinkedIn, using the hashtag #LAK21Accessibility. The workshop website can be found here. It provides details of the event and will also be used for an initial presentation of outcomes. Full outcomes of the workshop will be written up and submitted to the Journal of Learning Analytics.

4 REFERENCES

Clough, G., Adams, A., & Halford, E. (2017). Evidence Cafés and Practitioner Cafés supported by online resources: A route to innovative training in practice based approaches. *European Police Science and Research Bulletin: Special Conference Edition*, 3(3), 115-122.

Cooper, M., Ferguson, R., & Wolff, A. (2016). What can analytics contribute to accessibility in e-learning systems and to disabled students' learning? *LAK16*, Edinburgh, UK.

Lister, K., Coughlan, T., Iniesto, F., Freear, N., & Devine, P. (2020). Accessible conversational user interfaces: considerations for design. *Web for All 2020*, 20-21 April, Taipei, Taiwan.

Long, P., & Siemens, G. (2011). Penetrating the fog: analytics in learning and education. Educause Review, 46(5), 31-40.

Mamiseishvili, K., & Koch, L. C. (2012). Students with disabilities at 2-year institutions in the United States: factors related to success. *Community College Review*, 40(4), 320-339.

Nguyen, A., Gardner, L. A., & Sheridan, D. (2018). A framework for applying learning analytics in serious games for people with intellectual disabilities. *BJET*, 49(4). doi:10.1111/bjet.12625

Papathoma, T., Ferguson, R., Iniesto, F., Rets, I., Vogiatzis, D., & Murphy, V. (2020). Guidance on how Learning at Scale can be made more accessible. *Seventh ACM Conference on Learning@Scale*, 12-14 August, Atlanta, GA.

Tinto, V. (1997). Colleges as communities: taking research on student persistence seriously. *The Review of Higher Education*, 21(2), 167-177.

W3C. (2018). *Web Content Accessibility Guidelines (WCAG) 2.1*. Retrieved from http://www.w3.org/TR/WCAG/

Using Network Science in Learning Analytics: Building Bridges towards a Common Agenda

Oleksandra Poquet
C3L, University of South Australia
sasha.poquet@unisa.edu.au

Bodong Chen
University of Minnesota
chenbd@umn.edu

Mohammed Saqr
University of Eastern Finland
saqr@saqr.me

Tobias Hecking
German Aerospace Center
tobias.hecking@dlr.de

ABSTRACT: Interest in using networks in the analysis of digital data has long existed in learning analytics (LA). Applications of network science in our field are diverse. Some researchers analyze social settings in online discussions, knowledge building software, and group formation tools. Others use networked techniques to capture epistemic and cognitive processes. Networked approaches have been pioneered for psychometrics, for the analysis of time-series data, and for various types of clustering of relational observations. Finally, modelling of variables where networks are used as representations of causal relations is also gaining traction. Given the diversity of the thematic foci that researchers engage in when applying network science to learning analytics, this workshop aims to identify common challenges experienced through the use of network science methodologies. The workshop will invite researchers working in the area to share their work and reflect on common challenges. We envision themes of causality, linkage between micro- and macro-processes, use of time and space, elements of generalizability and validity to surface in the group discussions. The workshop aims to gather LA scholars to collectively build a solid foundation of advanced network modeling of learning data and shape strategies of future work in this important sub-field of LA.

Keywords: learning analytics, network science, common challenges

1 WORKSHOP ORGANIZERS

The workshop will be organized by four learning analytics researchers; two representing European institutions; one - North American, and one - Australasian. All of the organizers have been active in applying network science in their learning analytics work. They bridge different scholarly groups within the Society and diverse aspects of network science applications. The group has potential to bring together researchers working in networked learning, network science, learning sciences, educational research, computer science, social network analysis, knowledge building, and computer-supported collaborative learning. The organizers will leverage their scholarly networks to engage invited speakers for the workshop. The organizers have gone through workshop preparations last year and will build on lessons learnt.

2 WORKSHOP BACKGROUND

Social network analysis and its sister area of network science is widely used in learning analytics (LA). When positioned within a broader context, LA's focus on quantification of social interactions using digital data is not surprising. Early 2000s were characterized by the wider adoption of the web 2.0 in educational technology, and distance education pedagogies where these technologies were used, have always emphasized learner-to-learner interactions. Moreover, higher education literature referred to the outcomes of social interactions, such as social capital and the sense of belonging, as essential for student retention (Dawson, 2006). In K-12 schools, network analysis has been used to examine racial segregation in schools and further seek ways to support academic success of students from disadvantaged community (Zirkel, 2004; Farmer-Hinton, 2008). As a result, analyzing networks has been applied in a range of contexts: university online courses, MOOCs, social text- and video-annotation scenarios, as well as informal learning settings (Hoppe, 2017). Capturing learner interactions as network representations also potentially could be used for reflection and visualization of social dynamics in online course forums in LA dashboards.

Examples of empirical work analyzing social dynamics in socio-technical networks are diverse, including identification of network structures in different technological and pedagogical contexts; inquiry into the relationship of individual SNA metrics with performance and learning-related outcomes; clustering learners based on relational activities; examining learner positioning in relation to other indicators; detection of learner communities; modelling processes generating online learner networks; demonstrating group-level epistemic views, among others. However, analysis of social dynamics in socio-technical environments is not the only area of application for network science in learning analytics. As sophistication of computational approaches and collected data grew, so did the use of networks' scientific methodologies. The problems that can be studied using a network lens are as diverse as the contexts where they are applied, and far from uniform. Recent adoption of epistemic network analysis and growth of mixed methods in networked methods in one of the EARLI SIGs is just one example. Researchers in NetSci community also use network-related methods to model individual cognition, mental scheme networks, and language networks using common lexicon (Siew, et al., 2019). These network techniques broadly capture epistemic and cognitive processes for collective and individual systems, groups and individuals. Networked approaches have recently been adopted for the analysis of fine-grained time series data, and pioneered in psychometrics, as models combining various variables contributing to individual states (Marsmann et al., 2018). Finally, modelling of variables, using networks as representations of causal models is also gaining traction.

Given the diversity of thematic foci that researchers engage in when applying network science to learning analytics, this workshop aims to help researchers identify common challenges in their work, through the use of network science methodologies. The workshop will invite researchers working in the area to both share their work and reflect on common challenges. We envision themes of causality, linkage between micro- and macro-processes, use of time and space, as well as elements of generalizability and validity to surface in the group discussions. We envision this conversation to broadly evolve around best practices for operationalization of models that apply network scientific techniques and common research questions that fundamentally build on the complexity science approaches to modelling various systems (e.g. individual cognition, group cognition, epistemic structure, system dynamics, etc.). Another potential outcome is the discussion around the standards in reporting network analysis outcomes for easier replication.

3 ORGANIZATIONAL DETAILS OF PROPOSED EVENT

Type of event: Open symposium and Interactive small group discussions
Proposed schedule and duration: Half-day
Type of participation: Open participation, meaning any interested delegate may register to attend. To present, participants need to submit proposals for consideration. Proceedings for the workshop will be organized to encourage quality contributions.

The workshop activities that participants should expect: The workshop will invite participant contributions the exemplify current research or summarize past work into a coherent body of knowledge. We will ask participants to present their contributions in short talks (7-10 min., open format). 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. In their talks, the participants will be asked to reflect on current challenges inherent to their work that they would like to be addressed in the future. Two discussants will comment on work presented, synthesizing it. We expect the symposium (participant presentations and discussant comments) to take place for two hours. In the remaining hour, participants will be invited to work in small groups to discuss next steps that can be taken to address some of these challenges through collective work. At the end of the workshop, groups will be asked to share their insights. The event will take place via established video conferencing tools but will be scaffolded through distributed facilitated discussions in break-out rooms and supported by structured tasks. We hope to focus on common challenges, to provide participants with a shared space where their diverse thematic foci can be brought together.

Expected participant numbers and planned dissemination activities to recruit attendants: up to 20 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 reach out to the Network Science (NetSci) and EARLI network analysis group, as well as 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.

4 WORKSHOP OBJECTIVES OR INTENDED OUTCOMES

The workshop objectives are three-fold: to explore the application of advanced network analysis and modeling to learning data; to engage in discussion around the use of network science; and to identify common pain points that we collectively can work on. We hope to identify common areas that need improvement (e.g. framework for reporting results in network studies) that can align research efforts. This is a researcher-oriented community-building workshop, hence, the underlying goal is to enable space for researchers using network science to share and engage with one another, as a sub-community leading the development of this area. Submissions for the workshop will include short empirical papers, conceptual papers, and work-in-progress. They will be peer-reviewed. Accepted workshop papers will be published on the workshop website and in a joint LAK Companion Proceedings. The accepted papers will also be submitted for CEUR workshop proceedings for consideration. CEUR proceedings are SCOPUS indexed and allow for wider dissemination of the contributed work.

5 STRUCTURE AND CONTENTS OF THE WORKSHOP WEBSITE

We will set the workshop website, adding to the work submitted last year, using the Github Pages. The website will contain the workshop program, link to the workshop proceedings, and notes from group conversations, if the participants contribute them.

REFERENCES

Dawson, S. (2006). A study of the relationship between student communication interaction and sense of community. The Internet and Higher Education, 9(3), 153-162. http://doi.org/10.1016/j.iheduc.2006.06.007

Farmer-Hinton, R. L. (2008). Social capital and college planning: Students of color using school networks for support and guidance. Education and Urban Society. http://doi.org/10.1177/0013124508321373

Hoppe, H. U. (2017). Computational methods for the analysis of learning and knowledge building communities. Handbook of learning analytics, 23-33.

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. http://doi.org/10.1080/00273171.2017.1379379

Zirkel, S. (2004). What will you think of me? Racial integration, peer relationships and achievement among white students and students of color. Journal of Social Issues, 60(1), 57–74. https://doi.org/10.1111/j.0022-4537.2004.00099.x

Siew, C. S., Wulff, D. U., Beckage, N. M., & Kenett, Y. N. (2019). Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. Complexity, 2019 https://doi.org/10.1155/2019/2108423

Advancing Social Influence Models in Learning Analytics

Joshua Rosenberg
University of Tennessee, Knoxville
jmrosenberg@utk.edu

Bret Staudt Willet
Michigan State University
bret.staudt.willet@gmail.com

ABSTRACT: Our goal is to contribute to the Learning Analytics & Knowledge conference workshop on Using Network Science in Learning Analytics by advancing the use of a particularly important, but not widely used network science technique: modeling influence. Influence, the process through which individuals affect one another, has long been a key construct in social network analysis, but these models are uncommon in learning analytics-driven uses of network approaches. In this paper, we review prior educational research using influence models, provide an example from our recent work, and articulate some future directions for the use of influence models. We conclude with a description of how this work can contribute to the conference workshop and a call to consider how influence can complement selection and other network techniques used in learning analytics research.

Keywords: social network analysis, social influence, social capital, exposure effects, social media

1 INTRODUCTION

Network analysis is a complex methodological and theoretical lens through which a range of learning-related constructs can be examined. This complexity extends to learning analytics-driven uses of the network concept. This complexity has numerous effects. First, studying networks can be both compelling and challenging. This is particularly true for the networks learning-analytics scholars study, such as networks evidenced through conversation threads in online courses (e.g., Chen et al., 2019). These online networks may differ in fundamental ways from face-to-face networks for which network analysis has more often been used, such as advice-seeking networks among teachers in the same school building (Frank et al., 2004). These differences mean that although some established methods can be used, others must be modified, and, in some cases, new techniques must be developed. Consequently, another key product of the complexity of network analysis is that associated methods are likewise complex. That is, a range of methodologies that can be—and have been—used to analyze networks. This is especially true in the new terrain of data accumulated by educational technologies and learning analytics platforms, as well as digital-trace data and metadata from social media platforms.

This proposal will contribute to the Learning Analytics & Knowledge conference workshop on Using Network Science in Learning Analytics by advancing the use of a particularly important, but not-widely-used network analytic technique: modeling social influence. Influence has long been a key construct in social network analysis (Frank, 1998). For instance, sociologists developing the social influence approach used statistical models to understand how social capital (i.e., resources inherent to and available through relationships) exerted its power (Bourdieu, 1980). In short, influence may be thought of in terms of how individuals affect one another (Frank, 1998).

Although social influence may seem to be an essential characteristic of network studies, a review of research on social network analysis in learning analytics reveals a strong preference for another type of network process: social selection. Selection models aim to understand who interacts—and potentially forms relationships—with whom (Fincham et al., 2018). These selection processes are contemporarily estimated using powerful extensions of inferential statistical techniques such as logistic regressions, Exponential Random Graph Models (ERGMs; e.g., Gašević et al., 2019).

Social influence is distinct from—but also complements—social selection; these processes likely exist in a reciprocal relationship (Frank & Fahrbach, 1999). Furthermore, much of the existing social network analysis literature draws from descriptive statistical and visual approaches to understand networks. Each of these methodologies contributes its own distinct understanding to the phenomenon of social networks. However, social influence is currently under-represented in the current buffet of methods.

Our central argument here is that influence models are especially valuable because they allow researchers to interrogate what is intuitively important about networks. That is, it may seem self-evident that social networks can influence actions, behavior, and learning. However, measuring these phenomena can be difficult without the aid of the rather advanced statistical techniques of influence models.

To advance the understanding and adoption of influence models in learning analytics, we offer four pieces in this proposal. First, we provide a review of prior research in education on the use of influence models to understand networks. Second, we illustrate the use of influence models in the context of a recent study that explored influence in the context of an informal, technology-based online community of science educators. Third, we constructively critique our past research and suggest ideas for future work, whether these are our own efforts or those of other learning analytics researchers. We specifically highlight influence models for the effect of relationships in a network, which we consider to be a core yet missing element of network analysis. Fourth, and finally, we conclude with a description of how we see this work as contributing to the aims of the workshop.

2 PRIOR RESEARCH INVOLVING INFLUENCE MODELS

The prior research that has utilized influence models has primarily done so in the context of studies of the face-to-face networks of educators, teacher leaders, and administrators. For example, Frank et al. (2004) examined how the use of innovative digital technologies, namely the use of computers for five specific educational goals and activities, were adopted by teachers throughout a district when teachers identified as leaders among their peers adopted and productively used the tool. They collected network data from all of the teachers in the district by asking them to nominate up to ten individuals who they go to for help. Then, they determined how much of the variability in teachers' use of computer technologies depended upon who they said they went to for help over the preceding year. Counter to prevailing trends in educational technology research that has focused upon individual characteristics (often psychological), Frank et al. found that more variance in computer use was explained by social capital measures—who teachers went to for help— than more traditional, psychologically-focused measures of teachers' value for computers. The authors interpreted that it was through social capital (and social relationships) that teachers were exposed to expertise in a meaningful, context: a relationship with a trusted peer.

Another, more recent, example was reported by Horn et al. (2020), who focused on the nature and effects of the discussions that teachers had in workgroups. Extending their own and others' work that examined not just that influence took place (e.g., Coburn et al., 2010), Horn et al. examined how influenced was a function of the depth of the conversations that took place among teachers when exposure to expertise might occur. In other words, whereas Frank et al. (2004) assumed that when teachers nominated others (i.e., those who they turned to for help) this help is provided, Horn et al. modeled the kinds of substantive discussions that took place among those with differing expertise. This latter study found that those teachers who regularly participated in rich discussions about (mathematical) content were more likely to teachers' developing greater expertise.

These prior studies and other research (e.g., Cannata et al., 2010, Frank et al., 2020; Reddy et al., 2017; Sun et al., 2014) demonstrate that social influence can account for a great deal of the variance in key outcomes. Our contention is that these examples, which sometimes frame influence in terms of "exposure" (to expertise; Frank et al., 2004) effects, prompt questions for learning analytics research, too. For instance, relevant questions may include whether social interactions that take place in digital contexts for educational purposes (e.g., for teachers or learners participating in online learning communities) really matter. If so, how do these interactions matter (e.g., social influence)? In the next section, we describe a recent study in which we attempted to understand whether, and how, involvement in a social-media-based community for science educators influenced participants' sustained involvement over time.

3 AN ILLUSTRATION: INFLUENCE WITHIN #NGSSCHAT

To illustrate a recent effort to model social influence, here we describe a project focusing on science educators' adoption of the Next Generation Science Standards (NGSS). Specifically, educators have connected and interacted through a synchronous Twitter chat (#NGSSchat) to form a social-media-based professional network used to discuss topics related to the current science standards (i.e., the NGSS) in the United States (Rosenberg et al., 2020). In this study, we used public data mining methods to access more than 7,000 #NGSSchat posts, by around 250 participants, to one of approximately 50 one-hour synchronous "chats" that took place over two years, from 2014-2016. During these chats, participants discussed topics ranging from how to effectively communicate with parents about the new science standards to interpreting and discussing the research that contributed to the new standards.

Our goals in studying #NGSSchat were to (a) describe the depth of conversations that took place, (b) understand who was selecting to interact with whom, and (c) determine to what extent someone's future participation in the network was a function of with whom they interacted. The first and second goals were important for determining whether this social media-based network fostered meaningful conversations. That is, we wanted to know whether #NGSSchat discussions were "balanced" in terms of an egalitarian mix of posts going between researchers and teachers (i.e., not merely from researchers to teachers) and detailed (i.e., not predominantly superficial posts). The third goal specifically pertains to influence. If #NGSSchat operated like the face-to-face networks described in the previous section, then we would hypothesize that some type of social influence was likely taking place. Furthermore, we would surmise that this social influence bolsters the Twitter #NGSSchat network in the eyes of science educators who might understandably be skeptical about the value of this community.

To model influence, we examined how participation in #NGSSchat across an entire year could explain the rate of participation in the following year. We used a general linear model (with a Poisson outcome distribution because the dependent variable was a count) to predict sustained participation. We operationalized sustained participation as the number of original tweets each individual sent to #NGSSchat in one academic year (2015) as a product term representing involvement in each of the types of conversations. This term was intended to capture not only how many conversations an individual participated in, but also how some conversations may matter more when sent by central individuals. Accordingly, calculating these terms involved determining the number of times every other individual interacted with each individual and then multiplying that number by a centrality measure (in-degree centrality). Thus, these terms were intended to account for participating in conversations in which one received replies from individuals central to the network. Finally, we summed these multiplied terms to create a total value, or *exposure* (to influential others) term, for each individual. Thus, our model was relatively simple: we predicted the number of posts individuals sent in the subsequent year on the basis of an *exposure term* reflecting their involvement in conversations with central (and therefore potentially influential) individuals. We also included a predictor term to take into account individuals' professional roles.

Our analysis showed that the degree of individuals' exposure to conversations (accounting for the centrality of conversation participants) was associated with greater sustained participation. Specifically, for every one-SD increase in the number of conversations in which an individual participated, individuals were likely to post 9-15 additional tweets (in log-odds units, β's = 1.43 – 1.83, $p < .001$) in the next year, accounting for individuals' professional roles. From this, we inferred that if involvement in transactional conversations can support individuals to feel like they belong, conversation exposure might be what causes individuals to choose to continue to participate in the network. In sum, our analysis of Twitter #NGSSchat showed that involvement in conversations (similar to Horn et al. [2020]) predicted later participation.

4 FUTURE DIRECTIONS FOR MODELING INFLUENCE

Throughout this paper, we have described how influence matters for key outcomes: learning, implementing new teaching practices and making progress toward educational improvement efforts. However, one critique of our illustration we wish to raise is related to whether our outcome (i.e., sustained participation) is actually important. We consider this critique as a worthy outcome for the same reasons that we think studying social influence in digital contexts is important: It can allow us to determine whether and how #NGSSchat interactions matter. In this way, studying an outcome endemic to the network, rather than one external to it (e.g., whether teachers implemented what they learned or discussed through #NGSSchat, as determined through an observational measure) leaves open the question of the role of #NGSSchat in the implementation of the new science standards.

The previously described study on #NGSSchat (Rosenberg et al., 2020) was not alone in utilizing an imperfect outcome, and other studies have also linked teachers' networks to the implementation of their classroom practice (Frank et al., 2020). Therefore, a key future direction for modeling influence will be to explore whether and how educators' and learners' participation in myriad networks impacts their learning, actions, and capabilities. The more interesting question is not whether networks impact these and other outcomes, but, rather, which outcomes networks affect, and how they do so. For example, given the lack of focus in social media research on new teachers' needs, we might investigate how new teachers' participation in informal online networks affects their teaching practice.

Another future direction concerns the makeup of exposure terms that are so critical to influence models. Network analysts face numerous decisions regarding how to construct these. For instance, exposure terms can be based on the number of interactions or whether or not individuals interacted. Moreover, the effects of interactions from different individuals can be calculated in different ways: some influence processes are cumulative, such as, for example, when individuals are exposed to expertise from varied individuals, whereas for others the average influence is more salient. Finally, the time period over which exposure is evaluated is critical, and, distinct from descriptive analyses, there must be a time period over which exposure takes place—and, so, multiple measures are needed. Similarly, there are nuances to sort out related to influence as the learning analytics field has begun to address in the context of tie formation—or selection (Fincham et al., 2018).

5 CONTRIBUTIONS TO THE WORKSHOP AND CONCLUSION

Because the Using Network Science in Learning Analytics workshop is intended to identify common challenges faced by network science scholars and to surface these challenges in a way that supports the advancement of this field, our presentation will address several of the detailed workshop themes, particularly causality, the linkage between micro- and macro-processes, and linkages across time. Our contribution is to broaden the kinds of network science techniques learning analytics scholars use. Social influence is a model for network processes that differ from selection models that predict tie formation and network structure (e.g., Fincham et al., 2018), and is quite different from descriptive analyses that compute individual- and network-wide statistics, or simply present network visualizations. Several specific ways that this work will add to the workshop is to prompt discussions of (a) what kinds of questions are suited to the use of influence models, (b) how influence is similar to and different from other approaches, especially selection model effects through ERGMs, and (c) what the relative absence of influence models in the literature suggests about potential gaps in the growing body of learning analytics research that utilizes network science techniques—and what addressing those gaps might yield.

REFERENCES

Bourdieu, P. (1980). Le capital social: notes provisoires. Actes de la Recherche en Sciences Sociales, 31(1), 2–3.

Cannata, M., McCrory, R., Sykes, G., Anagnostopoulos, D., & Frank, K. A. (2010). Exploring the influence of National Board Certified Teachers in their schools and beyond. Educational Administration Quarterly, 46(4), 463-490.

Chen, B., & Huang, T. (2019). It is about timing: Network prestige in asynchronous online discussions. Journal of Computer Assisted Learning, 35(4), 503-515.

Coburn, C. E., Choi, L., & Mata, W. (2010). “I would go to her because her mind is math”: Network formation in the context of a district-based mathematics reform. In A. J. Daly (Ed.), Social network theory and educational change (pp. 33–50). Harvard Education Press.

Fincham, E., Gašević, D., & Pardo, A. (2018). From social ties to network processes: Do tie definitions matter? Journal of Learning Analytics, 5(2), 9-28.

Frank, K. A. (1998). Chapter 5: Quantitative methods for studying social context in multilevels and through interpersonal relations. Review of Research in Education, 23(1), 171–216.

Frank, K. A., & Fahrbach, K. (1999). Organization culture as a complex System: Balance and information in models of influence and selection. Organization Science, 10, 253–77.

Frank, K. A., Zhao, Y., & Borman, K. (2004). Social capital and the diffusion of innovations within organizations: The case of computer technology in schools. Sociology of Education, 77(2), 148-171.

Frank, K. A., Kim, J., Salloum, S. J., Bieda, K. N., & Youngs, P. (2020). From interpretation to instructional practice: A network study of early-career teachers’ sensemaking in the era of accountability pressures and Common Core state standards. American Educational Research Journal, 57(6), 2293-2338.

Horn, I., Garner, B., Chen, I. C., & Frank, K. A. (2020). Seeing colleagues as learning resources: The influence of mathematics teacher meetings on advice-seeking social networks. AERA Open, 6(2).

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.

Rosenberg, J. M., Reid, J., Dyer, E., Koehler, M. J., Fischer, C., & McKenna, T. J. (2020). Idle chatter or compelling conversation? The potential of the social media-based #NGSSchat network as a support for science education reform efforts. Journal of Research in Science Teaching, 57(9), 1322-1355. https://onlinelibrary.wiley.com/doi/10.1002/tea.21660

Sun, M., Wilhelm, A. G., Larson, C. J., & Frank, K. A. (2014). Exploring colleagues’ professional influence on mathematics teachers’ learning. Teachers College Record, 116(6), 1-30.

Modelling Network Dynamics in Social Annotation

Bodong Chen, Basel Hussein
University of Minnesota
chenbd@umn.edu, bhussein@umn.edu

Oleksandra Poquet
University of South Australia
sspoquet@gmail.com

ABSTRACT: Web annotation technology is used in education to facilitate individual learning and social interaction. Departing from a conceptual exploration of social interaction in web annotation as a mediated process, as well as a dissatisfaction with analytical methods applied to web annotation data, we analyzed student interaction data from a web annotation environment following the Relational Event Modelling approach. Included in our modelling were various annotation attributes, a contextual factor of student groups, and several social and spatiotemporal factors related to network formation. Results indicated that longer annotations were slightly more likely to attract replies, students in the same project group were not more likely to engage with each other, and several network factors such as student activity, reciprocity, annotation popularity, and annotation location played important roles in interaction dynamics. This study contributes empirical insights into web annotation and calls for future work to investigate mediated social interaction as a dynamic network phenomenon.

Keywords: web annotation, network analysis, collaborative learning, digital learning

1 INTRODUCTION

Computer-mediated communication tools (Jonassen et al., 1995) such as asynchronous online discussions and social network sites are often used to foster learner engagement and social interaction. Web annotation is one such tool for computer-mediated communication and learning. In a nutshell, a web annotation tool allows the user to highlight a target in a web document and post an annotation referring to that target (Sanderson, Ciccarese, & Young, 2017). The annotations make student thinking visible and encourages learners to interact with one another. This type of learning technology is not only well suited for active learning but could intensify the social nature of learning leading to improved motivation and meaningful social participation. Although the use of web annotation is widespread in education, research investigating spatial-temporal dynamics of individual participation and emergent social interaction in annotation environments is rare. A recent review on web annotation identified different ways to use it for learning (Zhu et al., 2020). However, previous studies do not attempt to reveal mechanisms driving social interaction in web annotation and offer limited practical guidance on how to facilitate student interaction.

This study aims to bridge this gap by examining social interaction in a web annotation environment used in an online class. The study had an overarching research question: *Which dynamics describe the formation of social interactions from individual annotation behaviors in web annotation?* In answering the question, we explicitly considered elements of course design, technology environment, and the spatiotemporal process of engaging with the environment. Following the Relational Event Modelling framework for the modelling of dynamic networks (Butts, 2008), we hypothesized that social interaction in the annotation environment was driven by three types of factors: *actor attributes* such as a learner's personal background, knowledge, and dispositions, *historical relational events* such as previous behaviors, and *exogenous contextual factors* such as being assigned to a group to collaborate on a course project. This study makes a major step in modelling mechanisms driving peer interaction within spatiotemporal and pedagogical constraints.

2 RELATED WORK

Due to its unique affordances, web annotation can be used to support social reading, group sense-making, knowledge construction, and community building (Chen, 2019; Kalir, 2020; Marshall, 1997). For example, *Hypothes.is*, a web annotation tool, creates layers of conversation on top of web documents. A student can highlight a piece of text and make an annotation, which can be responded to by other students in this group. Research shows such social annotation can support richer communication than other tools such as newsgroup and discussion boards, mostly because the annotated documents provide a context for engaged conversations (Su et al., 2010; Zhu et al., 2020)

To advance analysis, and eventually, interventions in web annotation environments, research needs to consider simple mechanisms that move digital learning spaces from individual interaction with content to the formation of social structures. We argue that a methodological shift in the analysis of social web annotation is needed. First, actor–artifact relations need to be considered seriously (Hoppe, 2017). Since social interactions are mediated by the annotation artifact, artifacts can have their own properties conducive, or not so much, to attracting others to engage. Further, to understand dynamics of social interaction, quantitative analysts need to model relational events that reflect the process, rather than collapsing a series of events into relational states between actors.

This study provides an example for modelling social interaction in a web annotation environment. We explicitly include in the modelling spatial and content properties of the annotation artifacts, course design elements that may have influenced learner activity, and the temporal process of events that took place. Despite being a case study, our modelling approach offers a generalizable view on the individual to social dynamics in the web annotation environment.

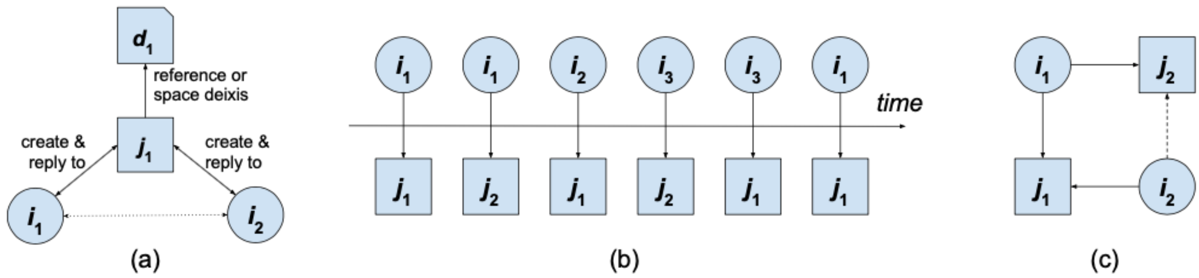


Figure 1. (a) Illustration of the mediated nature of social interaction in web annotation. An actor/author, i_1, creates an annotation, j_1, that references a document, d_1. Another actor, i_2, can indirectly communicate with i_1 by replying to i_2. (b) Illustration of the reply sequence. Actors/authors are represented as circles, annotations as squares. (c) The four-cycle closure is also illustrated, with the solid lines predicting the occurrence of the dotted line.

Following the Relational Event Modelling framework (Butts, 2008), the study examined mediated social interactions as relational events. In particular, we asked: Can we predict the occurrences of relational events between actors (e.g., learners) and annotations with node-level, social, spatial, or temporal factors in this context? Informed by prior work about social dynamics in online interactions (Chen & Poquet, 2020; Hecking, Chounta, & Hoppe, 2017; Poquet & Jovanovic, 2020; Vu, Pattison, Robins, 2015), we asked the following sub-questions:

- RQ1: To what extent did the attributes of students and web annotations contribute to mediated social interaction?
- RQ2: To what extent did the course-design factor of student grouping contribute to mediated social interaction?
- RQ3: To what extent did the endogenous network factors contribute to mediated social interaction?

3 METHODS

This study drew on a secondary dataset generated from an online course at a large public university in the United States. The course was a graduate-level seminar about Learning Analytics that involved 14 students. This course was designed to support collaborative knowledge building that demands extensive, emergent social interactions among students. A bulk of the class conversations took place on Hypothes.is, and students were asked to treat their annotations as a collective knowledge base for their course projects.

Using the Hypothes.is open API, we collected a total of 1,160 annotation events from the class, including 629 annotations and 531 replies. From the dataset, we generated an edge list for the two-mode, actor–annotation network; in each edge <*s*, *r*, *t*>, *s* stands for the sender—an actor/author, *r* the receiver—an annotation, and *t* the timestamp. Using the *rem* R package, this dataset was then structured as a discrete ordered sequence of relational events. Besides the network data, information about the annotations (e.g., location and word count) and the actors (e.g., project groups) was extracted.

In our empirical setting, reply actions can be generated only by an actor (i.e., a student) and directed toward an annotation. This is a one-plex, two-mode network, which has two types of nodes—actors and annotations, and one type of edge—reply. Different from a one-mode reply network that is typically constructed between participants, this two-mode network is based on an argument that a reply event from an actor to an annotation in the Hypothes.is environment also depends on traits of the annotation, even more so than characteristics of its author.

To address the specific research questions about factors driving social interaction in web annotation, we applied a novel network analysis approach named Relational Event Modelling (REM; Butts, 2008). Detailed explanations of REM can be found elsewhere (Butts, 2008). Briefly, a relational event is defined as a "discrete event generated by a social actor and directed toward one or more targets" (Butts, 2008). The central goal of REM is to "understand how past interactions effect the emergence of future interactions" (Butts, 2008), based on a set of derived statistics about: (a) endogenous network factors reflected in past relational events (e.g., frequency of previous actor–annotation events), (b) node attributes (e.g., an annotation's length and location, an actor's gender, activity level), and (c) exogenous contextual factors (e.g., trust, student groups).

In this study, we used the rem R package to compute a variety of network statistics (elaborated below) in response to our research questions. Using these statistics, we trained multiple relational event models following a forward selection strategy and evaluated model adequacy based on the AIC (Akaike information criterion) score (Butts, 2008). To model these computed statistics of relational events, we constructed a series of stratified Cox models using the *Survival* R package. We entered groups of variables in a stepwise manner and used the AIC score to evaluate whether the inclusion of new factors improved the models.

4 FINDINGS

On average, students wrote about 80 total posts consisting of 45 annotations and 35 replies during the 14-week semester. Temporally, we observed posting behaviors skewed (about 64%) toward the window between Sunday morning and Monday evening, right before the class meeting on Monday evenings.

Outputs of the relational event models are reported in Table 1. As indicated by the AIC score, adding the contextual factor of student grouping in Model 2 did not improve the model whereas the exogeneous network factors added in Model 3 greatly improved the model.

The first research question inquired about the extent to which artifact attributes effected learner interaction with them. Specifically, we examined if an annotations' location, length, and inclusion of a question mark were predictive of the likelihood of being replied. We observed from the models that the relative location of an annotation was not predictive of the likelihood of being replied, neither was the inclusion of a question mark in the annotation. The word count was positively associated with the likelihood of being replied, indicating longer annotations were more likely to receive replies. However, this effect was small.

The second research question asked whether contextual factors, related to pedagogical design in this class, effected mediated social interaction. Results from Model 2 and 3 showed that the project group homophily (i.e., learners being in the same group) was non-significant. That is, learners in the same group have not interacted with one another more than with other peers.

The third research question inquired about the role of exogeneous factors, related to emergent activity and dynamics between learners. To this end, we added four factors in Model 3. Results showed that the learner activity level, prior popularity of annotation, and "four-cycle" reciprocity, had contributed to more future interactions. Location homophily factor showed a negative effect. These findings indicated that in this social annotation context, the more active a student was, the more popular an annotation was, and the more likely they were going to be involved in the next reply event. The positive and significant effect of the "four cycle" showed that students have an inclination to collaborate with prior collaborators (see Figure 1(c)). However, the negative and significant effect of the location homophily showed that when an annotation receives a reply, other annotations near this annotation are likely to receive fewer replies.

Table 1. Relational event models of mediated social interaction in web annotation

	Model 1	Model 2	Model 3
Location of annotation	−.166	−.150	−.085
	(.169)	(.172)	(.180)
Question of annotation	.003	.002	.065
	(.108)	(.108)	(.112)
Length of annotation	.003	.003	**.003***
	(.001)	(.001)	(.002)
Project group homophily		.001	.004
		(.003)	(.003)
Actor activity			**.121*****
			(.017)
Location homophily			**−.010*****
			(.002)
Annotation popularity			**1.121*****
			(.075)
Reciprocity (four cycle)			**1.253*****
			(.160)
AIC	3677.903	3679.659	3269.597
R^2	.000	.000	.005
McFadden's pseudo R^2	.001	.001	.115

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

5 DISCUSSION AND CONCLUSIONS

The study was interested in examining mediated interaction and spatiotemporal dynamics in social web annotation. To foreground the mediated nature of social interaction in web annotation environments, we constructed a two-mode, bipartite network involving actors (students) and annotations. To examine network dynamics, we applied the Relational Event Modelling framework and modeled the impact of a number of factors on students' mediated social interaction.

This study also draws attention to the spatiotemporal properties of mediated social interaction in web annotation environments. The relational event modelling incorporates temporal information by design. Results showed network factors such as actor activity and annotation/artifact popularity played a role in the temporal evolution of the network. The addition of spatial information about annotations further allowed the investigation of temporal and spatial dynamics in tandem. Even though the annotation's spatial location was not predictive of network formation, we found a reply to an annotation would suffocate potential replies to nearby annotations. This mechanism might reflect the temporally condensed activities from individuals and temporally distanced activities among them (Chen & Huang, 2019). In other words, it is plausible in the web annotation context that some students would log on in a particular time when other students are unlikely to be active, scroll through a stack of peer annotations, and selectively reply to a few in different locations. The revealed spatiotemporal properties of social interaction in web annotation are worth considering if there is an interest in promoting peer interaction through instructional interventions.

In conclusion, this study contributes fresh insights into social interaction in web annotation, calls for attention to micro-level spatiotemporal patterns, and calls for future work to investigate mediated social interaction as a dynamic network phenomenon.

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. (2019). Designing for Networked Collaborative Discourse: An UnLMS Approach. *TechTrends*, *63*(2), 194–201. https://doi.org/10.1007/s11528-018-0284-7

Chen, B., & Huang, T. (2019). It is about timing: Network prestige in asynchronous online discussions. *Journal of Computer Assisted Learning*, 503–515. https://doi.org/10.1111/jcal.12355

Chen, B., & Poquet, O. (2020). Socio-temporal dynamics in peer interaction events. *In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, Frankfurt Germany, 203–208*. https://doi.org/10.1145/3375462.3375535

Hecking, T., Chounta, I. A., & Hoppe, H. U. (2017). Role Modelling in MOOC Discussion Forums. *Journal of Learning Analytics*, *4*(1), 85–116. https://doi.org/10.18608/jla.2017.41.6

Hoppe, H. U. (2017). Computational methods for the analysis of learning and knowledge building communities. In C. Lang, G. Siemens, A. Wise, & D. Gasevic (Eds.), *Handbook of learning analytics* (1st ed., pp. 23-33). Society for Learning Analytics Research (SoLAR). https://doi.org/10.18608/hla17.002

https://doi.org/10.1145/3375462.3375533

Jonassen, D., Davidson, M., Collins, M., Campbell, J., & Haag, B. B. (1995). Constructivism and computer-mediated communication in distance education. *American Journal of Distance Education*, *9*(2), 7–26. https://doi.org/10.1080/08923649509526885

Kalir, J. H. (2020). Social annotation enabling collaboration for open learning. *Distance Education*, *41*(2), 245–260. https://doi.org/10.1080/01587919.2020.1757413

Marshall, C. C. (1997). Annotation: from paper books to the digital library. *In Proceedings of the Second ACM International Conference on Digital Libraries - DL '97, New York, New York, USA, 131–140*. https://doi.org/10.1145/263690.263806

Poquet, O., & Jovanovic, J. (2020). Intergroup and interpersonal forum positioning in shared-thread and post-reply networks. *In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge, Frankfurt Germany, 187–196*.

Sanderson, R., Ciccarese, P., & Young, B. (2017). *Web annotation data model*. Technical Report, W3C.

Su, A. Y. S., Yang, S. J. H., Hwang, W.-Y., & Zhang, J. (2010). A Web 2.0-based collaborative annotation system for enhancing knowledge sharing in collaborative learning environments. *Computers & Education*, *55*(2), 752–766. https://doi.org/10.1016/j.compedu.2010.03.008

Vu, D., Pattison, P., & Robins, G. (2015). Relational event models for social learning in MOOCs. *Social Networks*, *43*, 121–135. https://doi.org/10.1016/j.socnet.2015.05.001

Zhu, X., Chen, B., Avadhanam, R. M., Shui, H., & Zhang, R. Z. (2020). Reading and connecting: using social annotation in online classes. *Information and Learning Sciences*, *121*(5/6), 261–271. https://doi.org/10.1108/ils-04-2020-0117

Idiographic Learning Analytics: A single student (N=1) approach using psychological networks

Mohammed Saqr
University of Eastern Finland
mohammed.saqr@uef.fi

Sonsoles López-Pernas
Universidad Politécnica de Madrid
sonsoles.lopez.pernas@upm.es

ABSTRACT: Recent findings in the field of learning analytics have brought to our attention that conclusions drawn from cross-sectional group-level data may not capture the dynamic processes that unfold within each individual learner. In this light, idiographic methods have started to gain grounds in many fields as a possible solution to examine students' behavior at the individual level by using several data points from each learner to create person-specific insights. In this study, we introduce such novel methods to the learning analytics field by exploring the possible potentials that one can gain from zooming in on the fine-grained dynamics of a single student. Specifically, we make use of Gaussian Graphical Models —an emerging trend in network science— to analyze a single student's dispositions and devise insights specific to him/her. The results of our study revealed that the student under examination may be in need to learn better self-regulation techniques regarding reflection and planning.

Keywords: Graphical Gaussian Models, Idiographic Learning Analytics, Network Science, Psychological Networks.

1 INTRODUCTION

The growing field of learning analytics (LA) has drawn the attention of academics, researchers, and administrators who aspire to understand and optimize teaching and learning (Siemens, 2013). Over ten years of findings have brought immense insights to our attention. One of the most important lessons that we have learned is that context matters: models obtained in one context are barely transferable to other contexts (Gašević et al., 2016). Researchers have failed to replicate the results of predictive models (e.g., for estimating student performance) across multiple learning settings due to the remarkable diversity in the data generated by students' learning activates, the obtained predictors, as well as the levels of statistical significance (Conijn et al., 2017; Dawson et al., 2019). These inconsistencies have made the efforts towards offering adaptive learning or personalizing support an arduous endeavor. Researchers have called for using the high resolution data generated by students to generate *personalized* insights (Winne et al., 2017). However, analyzing cross-sectional (i.e., group-level) data to generate *personalized* recommendations does not mean that each individual person will conform to the group average, and consequently, such insights generated by averaging over a group are hardly transmutable to every individual person (Fisher et al., 2018). Furthermore, cross-sectional group-level data fail to account for the dynamic processes (e.g., cognition and communication) that unfold within the individual. Obviously, a single cross-sectional timepoint is hardly useful to explain a dynamic phenomenon occurring over multiple time points (Beltz et al., 2016).

On this basis, idiographic methods have started to gain grounds as a possible solution to examine behavior at the individual level in other fields. Idiographic methods use several data points from an individual to create person-specific insights. Being derived on the person level, such analyses account for the individual factors while being able to explain dynamic phenomena (Epskamp, Waldorp, et al., 2018; Lamiell, 1981; Molenaar & Campbell, 2009). Winne et al. (2017) argued that high resolution data enable individual (i.e., idiographic) learner analytics, so that learners can gather own data and "interpret results to decide whether and how to adapt study tactics and learning strategies". Dawson et al. (2019) examined a large sample of students and tried early interventions aiming at prevention of dropouts. Their findings pointed to no effect on the retention outcome. The authors concluded that more data about individual differences are needed to better understand the retention process as well as to design relevant *personalized* interventions. A recent massive scale study that has examined a large sample of students (around 250,000) have found small benefit of a group-based behavioral intervention despite the massive dataset. Authors concluded that the field needs efficient interventions tailored to the *individual* and course context. Thus, education researchers need to explore such individual-based approaches (Beltz et al., 2016).

This study builds on the aforementioned insights and takes inspiration from the emerging fields of idiographic psychology and precision medicine, which have developed methods and standards for such methods of analysis (Beltz et al., 2016; Costantini et al., 2015; Molenaar & Campbell, 2009). In doing so, we explore the possible potentials that one can gain from zooming in on the fine-grained dynamics of a single student. We explore a person-specific data collection method as well as person-specific analysis and recommendations. Using data from a single student over 30 days, we analyze his/her dispositions and devise insights specific to him/her. Our approach is based on the emerging trend in network science, in particular, Gaussian Graphical Models (GGM) (Epskamp, Waldorp, et al., 2018; Saqr et al., 2021). Our research question is as follows: What insights can idiographic learning analytics reveal about students' self-regulation and learning dispositions?

2 BACKGROUND

2.1 The cognitive process as a networked system

Representing elements of the cognitive and social processes as a network is an established research method. Such representation has afforded researchers a way to visualize the structure of these processes to measure the magnitude of association between their elements, and to devise statistical indices that allow a precise interpretation of the resultant graphs (Dado & Bodemer, 2017). In education, research on networks spans three decades. Networks have been used to visualize the patterns of interactions in collaborative groups, to study the roles students play in the collaboration, to rank students' activities, or to predict performance to mention a few examples (Chen & Huang, 2019; Chen & Poquet, 2020; Halatchliyski et al., 2013; Saqr et al., 2019). While such methods have contributed enormously to our understanding of the learning process with their repertoire of powerful visualizations methods, there is a need for harnessing the power of other methods to extend our understanding different phenomena.

2.2 Gaussian Graphical Models

Recent advances in network sciences have led to the remarkable growth of probabilistic network models, often referred to as GGM (Epskamp, Waldorp, et al., 2018). GGM map the dynamic relationships between the elements of the cognitive or sociological phenomena we seek to understand as a complex system through the estimation of a network where the nodes are variables and the edges are the partial correlation coefficients between these variables (Artner et al., 2020; Borsboom, 2017; Epskamp, Waldorp, et al., 2018; Hamilton et al., 2019; Hevey, 2018). Similar to multiple regression, partial correlations estimate the correlations after controlling for all other variables in the network, thus eliminating the possible effect of confounding variables (Artner et al., 2020). This is particularly useful when there are multiple dependencies, i.e., consider an example when a researcher finds a positive correlation between coffee consumption and academic performance, such a correlation may simply be an unmeasured confounding factor (e.g., study time that leads to more coffee drinking). Thus, in GGM networks, two nodes are connected —if and only if— there is a covariance between these nodes that cannot be explained by any other variable in the network (Epskamp, Waldorp, et al., 2018; Hevey, 2018; Saqr et al., 2021). The resulting networks show only the significant relationships, the strength of such relationships, the sign (positive or negative), as well as the mediation pathways. Such rigorous network models offer “hypothesis generating structures, which may reflect potential causal effects to be further examined” (Hevey, 2018). As such, GGM offer several advantages that overcome the shortcomings of existing methods in terms of rigorous inferential statistics, ability to control for confounding factors, modelling the temporal evolution of the studied process. Moreover, there is a diverse and large community working on refining and improving GGM methods.

2.3 Graphical Vector Autoregression

An extension of GGM methods has allowed for the modeling of temporal processes, i.e., how a variable predicts another in the next time window. The abundance of intensive time-stamped data (time-series) has led to the existence of enough observations of individual subjects across short periods (e.g., experience sampling methods, observational data and physiological data), i.e., an individual can be studied as a unique case (N=1) (Epskamp, Waldorp, et al., 2018; Molenaar, 2004). Such time-series data are amenable to multivariate time-series analysis, commonly known as known as vector autoregression (VAR) (Epskamp, Waldorp, et al., 2018). VAR estimates a directed network (in contrast to undirected in GGM): the nodes are variables (e.g., motivation, behavior or attitude) and the link between them are temporal relationships (a variable predicts another in the next time window) (Epskamp, Waldorp, et al., 2018). This is commonly represented by drawing a directed arrow from the node that represents the variable (e.g., motivation) to the variable that it predicts in the next time window of measurement (e.g., engagement). An example is presented in Fig. 1, which shows a temporal network generated from a fictional individual dataset about hourly eating and exercise habits. The graph illustrates that running predicts rest thereafter and that comfort predicts eating (weak prediction, see the thin line). The loop around comfort means that comfort at one hour predicts that the person will be at comfort the next hour; probably breaking the eating habits may entail keeping occupied with activities. As shown, a temporal network predicts if a variable (an element of the studied phenomena) predicts another in the next time window. Such type of network is used to explain within-subject covariation or potential causal pathways.

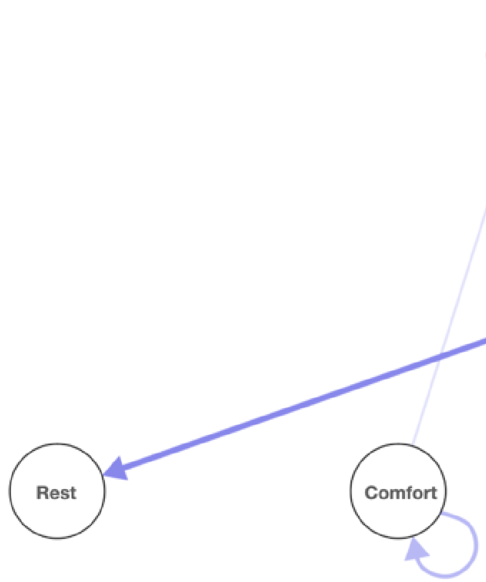


Figure 1: A fictional temporal network of four behaviors. The circles are variables. Blue lines are positive partial correlations. The thickness of the line is proportional to the magnitude of the correlation. The direction of the arrow points to the direction of the temporal correlation.

3 METHODS

The study included a single student who signed an informed consent for an anonymous version of the responses to be used for research purposes. The student was attending a course over a duration of a month. The student had to respond to ten questions representing common dispositions and self-regulation (SRL) that are commonly employed in learning analytics (Tempelaar, 2017; Tempelaar, Rienties, & Nguyen, 2018; Tempelaar, Rienties, Mittelmeier, et al., 2018). The questions covered the following constructs: Expectancy value (Vlu), Motivation (Mtv), Stress as negative affect (Str), Hope and enthusiasm as positive affect (Hop), SRL Planning (Pln), SRL Engagement with task (Tsk), SRL Reflection and evaluation (Rfc), External Regulation by assignments (Asg), Socializing (Soc), Challenging learning tasks (Chl).

The survey data was detrended using the method described in (Epskamp, van Borkulo, et al., 2018) to make the data close to stationary. Since our interest was to study the interplay between the student's different dispositions, we used the VAR model. VAR models have been established in the study of psychological phenomena, shedding light on the temporal progression, individual aspects and dynamics of psychological processes within individuals (Epskamp, van Borkulo, et al., 2018; Epskamp, Waldorp, et al., 2018; Fisher et al., 2017). To understand the sequential temporal dependencies, we created a temporal network by estimating a Graphical VAR model (Epskamp, van Borkulo, et al., 2018). The temporal network captures what will happen next as an effect of what is happening now (lag-1 or cross-lagged effects), e.g., if the person is motivated now, the person is going to work on the task on the next step. To account for multiple comparisons, the model was regularized using graphical least absolute shrinkage and selection operator (GLASSO). Using GLASSO algorithm for estimating GGM networks has been shown to retrieve the true structure of the network (Epskamp, van Borkulo, et al., 2018).

4 RESULTS

The results of the temporal network showed interesting results about the involved student (Fig. 2 and Table 1). After controlling for all other variables in the network, the positive affect (feeling hope) was the most predictive variable of engagement in a task in the next day, shown as a thick arrow between the *Hop* and *Tsk* nodes in Fig. 2 indicating the strong association. Motivation was also strongly predictive of engagement with the task after controlling for all other variables, i.e., independent of feeling hopeful, socializing, etc. The challenging nature of the task was also predictive of engagement for the student, as well as stress, indicating that a bit of a challenge may help some students engage and work on the learning activities. The expected value and relevance of the task was also predictive of the student's engagement with the task, emphasizing the need for creating more relevant and authentic learning tasks.

Working on the assignment was negatively predictive of engagement with learning tasks, as the student focused more on finishing the submissions. Such results also indicate that external regulation may be counterproductive for some students. Similarly, reflection was negatively predictive of engagement with the task the next day, which raises the question of the nature of reflection the student has. Planning was also weekly negatively associated with engagement with the task. These negative associations for assignment, reflection and planning are indicative of poor self-regulation practices by the student. In fact, the student had to repeat one of the assignments as it was not fulfilling the required guidelines and was incomplete. He also scored below the 50th percentile in the two most important course assignments. There is room for improvement here, by helping the student learn optimal self-regulation practices. There was a negative association between motivation and planning, while strong positive association with socialization. Stress and assignment negatively influenced each other: the more stress the student was under, the less he/she worked on the assignments, and the more work on assignments the less stressed was the student, as expected. The results are detailed in Table 1. Please note that, since partial correlations do control for other variables, their values are not to be interpreted in the same way, as they tend to be lower.

Table 1: Values of the VAR partial correlations

	Tsk	Vlu	Mtv	Str	Hop	Pln	Rfc	Asg	Soc	Chl
Tsk	0.00	-0.02	0.00	0.00	0.00	0.00	0.00	0.01	-0.01	0.00
Vlu	0.16	0.00	0.03	0.00	0.00	-0.02	0.00	0.04	0.10	0.08
Mtv	0.27	0.03	0.00	0.03	0.00	-0.07	0.03	0.00	0.28	0.00
Str	0.17	-0.07	0.00	0.00	0.05	-0.01	0.00	-0.08	0.12	0.03
Hop	0.29	0.06	0.00	0.00	0.00	-0.05	0.05	-0.02	0.06	0.13
Pln	-0.06	-0.07	0.00	-0.06	0.00	0.00	0.00	0.00	-0.04	-0.09
Rfc	-0.22	0.01	0.00	0.00	0.00	0.00	0.05	0.01	0.00	0.00
Asg	-0.17	0.05	0.00	-0.09	-0.03	0.00	0.05	0.18	-0.17	0.01
Soc	-0.02	0.04	-0.01	-0.03	-0.02	0.02	-0.01	0.00	0.01	0.03
Chl	0.20	0.00	0.09	0.00	0.04	0.00	-0.09	0.00	0.00	-0.06

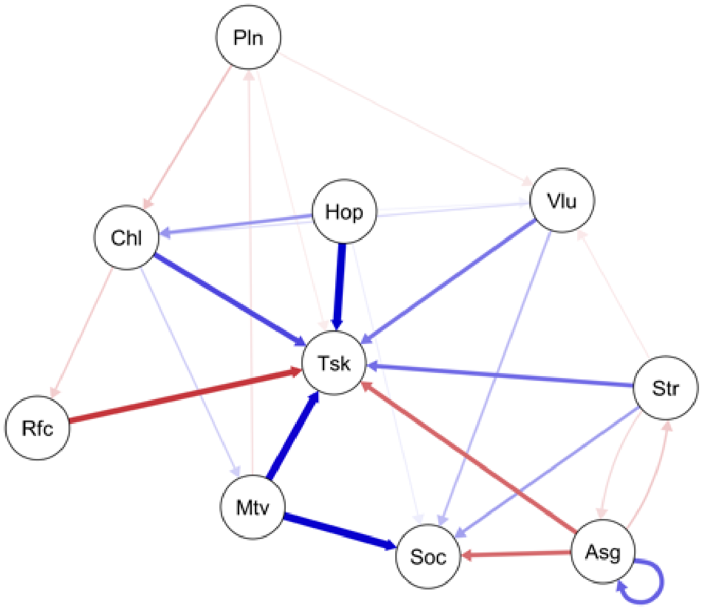


Figure 2: Temporal network for the student

5 DISCUSSION AND CONCLUSIONS

In this study, we have used psychological network methods in the form of GGM and graphical VARs to study a single student disposition during a course. Such idiographic method offers several advantages over cross-sectional group level analysis. Being focused on a single student, the insights generated are more relevant and actionable, i.e., precisely personalized, paving the path for precision education. These methods also offer several advantages regarding controlling for confounders, deleting spurious correlations and regularization which requires high magnitude significant correlation, offering a good level of rigorousness (Epskamp, van Borkulo, et al., 2018; Epskamp, Waldorp, et al., 2018; Fisher et al., 2017). The study has shown that the student under examination may be in need to learn better self-regulation techniques regarding reflection and planning based on his own responses. However, the value of such targeted intervention is yet to be investigated.

The implication of our study can be the applicability of the approach in several scenarios and contexts. Researchers who wish to apply personalized learning analytics can use such methods to design personalized intervention for their students. We believe there is an opportunity that may change the deserves attention and efforts from the research community to extend, improve and build on such methods. Our methods are not without limitations. The idea that the data have to be collected on a daily basis makes it sometimes difficult to collect data without some gaps, non-compliance, or missing values. The rate of data collection can be tricky: we have used a lag of a single day, but we do not know for sure if that lag was optimal. The timing of the data collection is another factor: whether data should be collected before or after the working day is still an open question. Similarly, how frequently data should be collected, what factors are to be included in the study, and how long we should collect the data are aspects in need of further investigation. The collection of data comes always with problems and risks of privacy and ethical concerns (Munoz-Arcentales et al., 2019; Saqr, 2017), in idiographic approach where much data is collected it can pose a risk which needs to be mitigated (López-Pernas & Saqr, 2021).

REFERENCES

Artner, R., Wellingerhof, P. P., Lafit, G., Loossens, T., Vanpaemel, W., & Tuerlinckx, F. (2020). The shape of partial correlation matrices. *Communications in Statistics - Theory and Methods*, *0*(0), 1–18. https://doi.org/10.1080/03610926.2020.1811338

Beltz, A. M., Wright, A. G. C., Sprague, B. N., & Molenaar, P. C. M. (2016). Bridging the Nomothetic and Idiographic Approaches to the Analysis of Clinical Data. *Assessment*, *23*(4), 447–458. https://doi.org/10.1177/1073191116648209

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, *16*(1), 5–13. https://doi.org/10.1002/wps.20375

Chen, B., & Huang, T. (2019). It is about timing: Network prestige in asynchronous online discussions. *Journal of Computer Assisted Learning*, *35*(4), 503–515. https://doi.org/10.1111/jcal.12355

Chen, B., & Poquet, O. (2020). Socio-temporal dynamics in peer interaction events. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 203–208. https://doi.org/10.1145/3375462.3375535

Conijn, R., Snijders, C., Kleingeld, A., & 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. https://doi.org/10.1109/TLT.2016.2616312

Costantini, G., Epskamp, S., Borsboom, D., Perugini, M., Mõttus, R., Waldorp, L. J., & Cramer, A. O. J. (2015). State of the aRt personality research: A tutorial on network analysis of personality data in R. *Journal of Research in Personality*, *54*, 13–29. https://doi.org/10.1016/j.jrp.2014.07.003

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. https://doi.org/10.1016/j.edurev.2017.08.005

Dawson, S., Joksimovic, S., Poquet, O., & Siemens, G. (2019). Increasing the Impact of Learning Analytics. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, *March*, 446–455. https://doi.org/10.1145/3303772.3303784

Epskamp, S., van Borkulo, C. D., van der Veen, D. C., Servaas, M. N., Isvoranu, A. M., Riese, H., & Cramer, A. O. J. (2018). Personalized Network Modeling in Psychopathology: The Importance of Contemporaneous and Temporal Connections. *Clinical Psychological Science*, *6*(3), 416–427. https://doi.org/10.1177/2167702617744325

Epskamp, S., Waldorp, L. J., Mõttus, R., & Borsboom, D. (2018). The Gaussian Graphical Model in Cross-Sectional and Time-Series Data. *Multivariate Behavioral Research*, *53*(4), 453–480. https://doi.org/10.1080/00273171.2018.1454823

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(27), E6106–E6115. https://doi.org/10.1073/pnas.1711978115

Fisher, A. J., Reeves, J. W., Lawyer, G., Medaglia, J. D., & Rubel, J. A. (2017). Exploring the idiographic dynamics of mood and anxiety via network analysis. *Journal of Abnormal Psychology*, *126*(8), 1044–1056. https://doi.org/10.1037/abn0000311

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002

Halatchliyski, I., Hecking, T., Göhnert, T., & Hoppe, H. U. (2013). Analyzing the Flow of Ideas and Profiles of Contributors in an Open Learning Community. *Proceedings of the Third International Conference on Learning Analytics and Knowledge - LAK '13*, *1*(2), 66–74. https://doi.org/10.1145/2460296.2460311

Hamilton, M., Clarke-Midura, J., Shumway, J. F., & Lee, V. R. (2019). An Emerging Technology Report on Computational Toys in Early Childhood. *Technology, Knowledge and Learning*. https://doi.org/10.1007/s10758-019-09423-8

Hevey, D. (2018). Network analysis: A brief overview and tutorial. *Health Psychology and Behavioral Medicine*, *6*(1), 301–328. https://doi.org/10.1080/21642850.2018.1521283

Lamiell, J. T. (1981). Toward an idiothetic psychology of personality. *American Psychologist*, *36*(3), 276–289. https://doi.org/10.1037/0003-066X.36.3.276

López-Pernas, S., & Saqr, M. (2021). Idiographic Learning Analytics: A Within-Person Ethical Perspective. *Companion Proceedings 11th International Conference on Learning Analytics & Knowledge (LAK21)*, in-press.

Molenaar, P. C. M. (2004). A Manifesto on Psychology as Idiographic Science: Bringing the Person Back Into Scientific Psychology, This Time Forever. *Measurement: Interdisciplinary Research & Perspective*, *2*(4), 201–218. https://doi.org/10.1207/s15366359mea0204_1

Molenaar, P. C. M., & Campbell, C. G. (2009). The New Person-Specific Paradigm in Psychology. *Current Directions in Psychological Science*, *18*(2), 112–117.

Munoz-Arcentales, A., López-Pernas, S., Pozo, A., Alonso, Á., Salvachúa, J., & Huecas, G. (2019). An Architecture for Providing Data Usage and Access Control in Data Sharing Ecosystems. *Proceedings of the 6th International Symposium on Emerging Information, Communication and Networks (EICN 2019)*, *160*, 590–597. https://doi.org/10.1016/j.procs.2019.11.042

Saqr, M. (2017). Big data and the emerging ethical challenges. *International Journal of Health Sciences*, *11*(4), 1–2.

Saqr, M., Nouri, J., & Fors, U. (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.

Saqr, M., Viberg, O., & Peeters, W. (2021). Using Psychological Networks to Reveal the Interplay between Foreign Language Students' Self-Regulated Learning Tactics. *Proceedings of the 2020 STELLA Symposium*, 12–23.

Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, *57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Tempelaar, D. (2017). How Dispositional learning analytics helps understanding the worked-example principle. *Proceedings 14th International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2017)*, 117–124.

Tempelaar, D., Rienties, B., Mittelmeier, J., & Nguyen, Q. (2018). Student profiling in a dispositional learning analytics application using formative assessment. *Computers in Human Behavior*, *78*, 408–420. https://doi.org/10.1016/j.chb.2017.08.010

Tempelaar, D., Rienties, B., & Nguyen, Q. (2018). Investigating learning strategies in a dispositional learning analytics context: The case of worked examples. *ACM International Conference Proceeding Series*, 201–205. https://doi.org/10.1145/3170358.3170385

Winne, P. H., Nesbit, J. C., & Popowich, F. (2017). nStudy: A System for Researching Information Problem Solving. *Technology, Knowledge and Learning*, *22*(3), 369–376. https://doi.org/10.1007/s10758-017-9327-y

Construction of Weighted Course Co-Enrollment Network

XunFei Li
University of California, Irvine
xunfeil@uci.edu

Renzhe Yu
University of California, Irvine
renzhey@uci.edu

ABSTRACT: Recent years have seen the increasing availability of campus administrative data and learning management system (LMS) data, researchers get the opportunity to explore how the co-presence relational data, rather than self-report friendship relational data affects students' educational outcomes. Social network analysis is one of the most used methods to study relational data, which can provide abundant information regarding how students are connected through the course co-enrollment network, as well as the effects of student's network property on their educational outcomes. This study intends to explore how to construct the course co-enrollment network using administrative data from a large public university in the United States and to explore if we observe different relation between network indices and students' academic performance between a more accurately defined and a simple course co-enrollment network. Furthermore, by comparing the network autocorrelation models between a simple course co-enrollment network built without edge weight difference and the weighted edge course co-enrollment network, we can explore if peers from the more heavily weighted classes are more influential than peers from the less weighted classes.

Keywords: Social network analysis, course co-enrollment, co-presence network, peer effect

1 INTRODUCTION

Course-taking experience plays a critical role in undergraduate students' college life. Peers in courses are one of the most important parts of students' course-taking experience. Not only the general course peers' composition (such as gender ratio, ethnicity composition) serve as a crucial social context of students' college experience, but the potential opportunity of direct (such as group work) and indirect interactions (such as presentations) are embedded among peers in the same course (Eckles & Stradley, 2012).

Although peer effects have been studied in educational contexts for a long time, most of the research focused on the effect of students' direct friends or the effect of student's roommates on their educational behaviors or performance (Biancani & McFarland, 2013). Recent years have seen the increasing availability of campus administrative data and learning management system (LMS) data, researchers get the opportunity to explore how the co-presence relationship, rather than self-report friendship affects students' educational outcomes. Social network analysis is one of the most used methods to study relational data, which can provide abundant information regarding how students are connected through the course co-enrollment network, as well as the effects of student's network property on their educational outcomes.

Studies applying social network analysis on course co-enrollment and co-presence networks find that network statistics such as degree and density contribute to explaining students' educational outcomes (Fincham et al., 2018; Israel et al., 2020; Weeden & Cornwell, 2020). However, little effort has been put into constructing the course co-enrollment network accurately based on the course level information. The strength of connections between students in different courses varies according to various factors such as course type, course format, course meeting schedule, etc. The relationship between network statistics and educational outcomes is affected by how the network is constructed. This study intends to study how to precisely construct the course co-enrollment network using administrative data from a large public university in the United States, and weighted network better approaches the relationship between network statistics and students' educational outcomes and the social influence within the network.

2 REVIEW OF LITERATURE

2.1 Network Study in Educational Field

Social network analysis (SNA) has been used in studying learning contexts for a long time (Biancani & McFarland, 2013). Traditionally, students' friendship network and students' roommate network have gained the most attention for examining how significant others' preference and selection affect students' educational performance and behaviors. SNA has also been applied to rich discussion forum data to understand how students interact with each other through discussion threads (Fincham et al., 2018). As various campus data (such as administrative data, WiFi log data) becoming increasingly available, a handful of studies in the field of higher education use course co-enrollment and co-presence network to study how network structure relate to students' connections and behaviors (Fincham et al., 2018; Nguyen et al., 2020).

2.2 Course Co-Enrollment Network

Course co-enrollment network is one of the most at scale and available data sources that capture the important relational data from students and their course peers during students' college time. However, few studies explore the effect of co-presence and students' peer influence through courses using course co-enrollment data. The challenge of using course co-enrollment data is how to understand and model the effects of students' co-presence/exposure and peer influence in a more accurate way. Students enrolling in the same course do not necessarily mean they have sufficient and equal exposure to each other.

2.3 Network Construction in Educational Field

Different network constructions represent researchers' understanding of the relational network data: in school friendship networks, the formation of a tie depends on students' self-report data that specifying their best friends, which assumes students' best friends have significant effects on them. Ties could also be constructed based on students' direct interactions. The network constructed with discussion forum data usually defines tie as students' direct interaction: responding posts. Network focus on small groups of people such as study group or orientation group's effect also assumes students affect each other through interactions. Co-presence networks define tie as students' co-presence in the same space during the same time, such as network constructed based on students WiFi log data, course co-enrollment, campus activities participation (Eckles & Stradley, 2012).

In respect of course co-enrollment network, it can be constructed as a two-mode course-student network or can be projected as two one-mode networks separately: student-student network and course-course network. Network structure and tie definition could be affected by time span and nodes inclusion of the course co-enrollment network based on research purposes (Gardner et al., 2018; Israel et al., 2020; K. A. Weeden & Cornwell, 2020). Weeden and Cornwell (2020), and Israel et al. (2020) both define the edge as student's co-enrollment in a class, but they include different student groups and time span in their networks. Weeden and Cornwell (2020) construct a two-mode course co-enrollment network with Cornell University's single term's transcript data. Undergraduate, graduate and professional master students are connected to each other if they are in the same class in that term, not connected if they are not in the same classes. All the ties are treated equally. Israel et al. (2020) project a one-mode course network and a one-mode student network from the full two-mode co-enrollment network, which is based on one cohort of students' six years of course-taking data. Students form tie with other students if they ever enrolled in the same class six years after they enrolled, edges are weighted through the total number of co-enrolled courses. Students outside of the cohort were not included in the networks. Gardner et al. (2018) use ten years of undergraduate courses taking the record to build the network and further specify different edges through link attributes, which change according to co-enrolled peers' attributes.

3 RESEARCH QUESTIONS

This study intends to construct students' course co-enrollment network with course-level information from university administrative data. The ties between students who co-enrolled in the same courses would be weighted by detailed course information such as course type, course size, and course meeting schedule. The assumption behind this construction is that the strength and the effect of peers' co-presence on students are not the same in different courses setting. Students may have more in-depth connections in small seminars than in large lectures, in classes with more frequent meeting schedules than courses with less opportunity to meet. This course-relevant information would affect the strength of students' connection through the course co-enrollment network.

Research question 1:

How to construct accurate course co-enrollment networks informed by campus administrative data which includes students' course enrollment, course-level information, and student-level information?

To further validate if a more accurate course co-enrollment network can represent students' connection built through course co-enrollment more effectively, this study also seeks to examine how students' academic performance correlated with each other in the course co-enrollment network through network autocorrelation models. By comparing the network autocorrelation network between a simple course co-enrollment network built without weight difference and the weighted course co-enrollment network, we can explore if peers in the classes with heavier weight are more relevant than peers from less weighted classes.

Research question 2:

Do we observe different relation between network indices and students' academic performance between a weighted network and a simple course co-enrollment network?

Research question 3:

Do we observe stronger social influence through the autocorrelation model in students' academic performance between students who share heavily weighed tie than between students who share light-weighted tie through the course co-enrollment network built by more informative courses' information? Does the weighted network fit the autocorrelation model better than a simple non-weighted course co-enrollment network?

4 METHODS AND PROPOSED ANALYSES

4.1 Data Sources

The data used in constructing the course co-enrollment network come from the administrative data from a large public university in the United States. The administrative data includes students' transcript data, students' courses taking trajectories, and courses relevant information of undergraduate students who enrolled in UCI from 2008 to 2020. The student population and course settings in this data allow us to observe the overall undergraduate students' course co-enrollment network with a representative perspective. A large public university would include most of the common majors and schools with students who come from different family backgrounds. The detailed student by term by course data with students' grade and course relevant information enable this study to see which students took the same courses and then construct their tie with their classmates. The course information provides us rich information to weigh the ties between students according to course features. The data used in this study's analysis only include data from 2015 to 2020 to follow the complete college experience of students from the 2015 cohort and 2016 cohort. The course only includes one student or students who did not complete the course are excluded from the analysis.

4.2 Course Co-Enrollment Network Construction

The course co-enrollment network is constructed as a one-mode network that each node represents one student (Zhou et al., 2007). Students have ties with other students if they enrolled and completed the same class. The network is an m*m matrix that m equal to the total number of students in that term excluding students who were only in courses with only one student or students who failed all classes.

Each cell in the matrix presents the weight of the tie of row m student and column n student. If they went to and completed the same class then their cell would be filled with 1 instead of 0. If row m student and column n student enrolled and completed more than one class, the cell would be filled with the total overlapping courses they had.

Weighted Ties

The existence of ties between students depends on whether they enrolled and completed the same courses, but not all the ties have equal values. Informed by the course relevant information, we calculate the edge weight through the combination of courses' information. Certain courses' features accord larger weight to ties generated in those courses. The course features we are using include Course types that include lecture, seminar, lab, and discussion. Different types would be assigned to different weights based on the chance of interaction students generally have in this course type, from the most heavily weighted course type to the least weighted course type: seminar, discussion, lab, lecture; Course schedule which includes the meeting times of the course. The course would be heavily weighted based on the frequency of the meeting schedule. Courses that meet more are weighted heavier than courses that meet less (Srinivasan et al., 2006); Course size which represents the total number of students in that course session. Smaller courses are weighted heavily because the chance of interaction between students is higher in smaller courses than larger courses; the Courses level captures the upper-division courses and the lower division courses. Upper-division courses are weighted heavier than lower-division courses since they generally ask for more engagement from students; Courses location and physical environment could also be included in the weight generating formula. Courses located in a larger lecture hall may provide less opportunity for students to interact with each other.

4.3 Network Autocorrelation Model

Network autocorrelation enables us to analyze the social influence process among people in an interdependent network (Leenders, 2002). In the autocorrelation model, ego's endogenous outcome variable is not only affected by the ego's own covariates but also affected by other alters in the same network with the ego. The strength of alters' influence is determined by the weight matrix in the autocorrelation model. In this study, students' term GPA (or cumulative) would be the endogenous outcome variable, the covariates include students' previous cumulative GPA, demographic characteristics (gender, race, first-generation college student status, low-income status).

The weight matrix is informed by the weighted ties we calculate based on course information, which could capture more accurate strength of students' influence to each other through courses co-presence than simply treat all the peers in co-enrolled courses equally. By using the autocorrelation model, we can test how peers in course co-enrollment network influence each other through different specifications of the weight of their ties with each other.

4.4 Discussion

The limitation of this study is the internal relationship and mechanism between students' co-presence and peer effect is still not clear. The precision of the course co-enrollment construction could help us to examine and compare if the strength of the co-presence network affects the relationship between network statistics and educational outcome, and the social influence within the network. This study could provide insight into how students in university connect to each other with different strength by different course settings, which could help policymaker in the field of higher education to better understand students' college experience through course taking and to explore course setting policy to help students connect and interact with students in a more meaningful way.

REFERENCES

Biancani, S., & McFarland, D. A. (2013). Social Networks Research in Higher Education. In Higher Education: Handbook of Theory and Research (Vol. 28, pp. 151–215). https://doi.org/10.1007/978-94-007-5836-0

Eckles, J. E., & Stradley, E. G. (2012). A social network analysis of student retention using archival data. Social Psychology of Education, 15(2), 165–180. https://doi.org/10.1007/s11218-011-9173-z

Fincham, E., Gašević, D., & Pardo, A. (2018). From Social Ties to Network Processes: Do Tie Definitions Matter? Journal of Learning Analytics, 5(2), 9–28. https://doi.org/10.18608/jla.2018.52.2

Israel, U., Koester, B. P., & McKay, T. A. (2020). Campus Connections: Student and Course Networks in Higher Education. Innovative Higher Education, 45(2), 135–151. https://doi.org/10.1007/s10755-019-09497-3

Leenders, R. T. A. J. (2002). Modeling social influence through network autocorrelation: Constructing the weight matrix. Social Networks. https://doi.org/10.1016/S0378-8733(01)00049-1

Nguyen, Q., Poquet, O., Brooks, C., & Li, W. (2020). Exploring homophily in demographics and academic performance using spatial-temporal student networks. Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020), 194–201.

Srinivasan, V., Motani, M., & Wei, T. O. (2006). Analysis and implications of student contact patterns derived from campus schedules. Proceedings of the 12th Annual International Conference on Mobile Computing and Networking (MobiCom ’06), 86–97. https://doi.org/10.1145/1161089.1161100

Weeden, K. A., & Cornwell, B. (2020). The Small-World Network of College Classes: Implications for Epidemic Spread on a University Campus. Sociological Science, 7, 222–241. https://doi.org/10.15195/V7.A9

Zhou, T., Ren, J., Medo, M., & Zhang, Y.-C. (2007). How to project a bipartite network? 1–7. https://doi.org/10.1103/PhysRevE.76.046115

Bayesian Knowledge Tracing with Python for Researchers and Practitioners

Zachary A. Pardos
University of California, Berkeley
pardos@berkeley.edu

Frederic Wang
University of California, Berkeley
fredwang@berkeley.edu

Anirudhan Badrinath
University of California, Berkeley
abadrinath@berkeley.edu

Cristian Garay
Microsoft
cgaray@gmail.com

ABSTRACT: Bayesian Knowledge Tracing (BKT) is a common statistical model used in intelligent tutoring systems to help adapt material by estimating when a student has mastered a skill. While the research community around BKT has been active, there has been a high barrier to entry given the lack of accessible software libraries. This tutorial will introduce participants to the first BKT library for Python, a computationally efficient implementation that allows for easy replication of many model variants from the literature. The tutorial will consist of 30 minutes of lecture and 2 hours of notebook-based, hands-on tutorial activities and group work.

Keywords: Bayesian Knowledge Tracing; Intelligent Tutoring Systems; Python Library

1. WORKSHOP DETAILS

1.1. Agenda

1.1.1. Presentation

(5 minutes) Introduction, History of BKT and knowledge tracing algorithms
(10 minutes) Fundamentals of BKT involving the base equations and HMMs, including demonstrations to how the algorithm works using simple examples of a few responses
(5 minutes) History of BKT software, including examples such as BNT and xBKT
(5 minutes) Advantages of using pyBKT over other software (computational efficiency, support of model variants, bug fixes with regards to xBKT, ease of use/accessibility using the fitting, cross validation and evaluation methods).(15 minutes) Demonstration of pyBKT, starting from simple synthetic data examples on the basic BKT model and then gradually working up to using cross validation on real data from ASSISTments and Cognitive Tutor data sets

1.1.2. Introductory Hands-On Problems

Have the attendees work through an introductory lab for pyBKT, where they will go through simple examples to get a feel for the software as well as answer questions regarding BKT. Examples will include edge cases uses of BKT to deepen understanding, application of cross-validation to explore the best model variant for a specific dataset and scenario, and other applications of BKT to real-world use cases. During this time attendees are free to ask questions and work with each other.

1.1.3. Informal Competition

During this time, we will host a friendly competition. We will give attendees a dataset of student responses, and attendees will be tasked with manipulating the input data and choosing both parameters and the model (possibly ones that they create themselves) to best predict student responses. Performance evaluation will be based on a combination of cross-validated accuracy, RMSE, and AUC.

1.2. Learning Objective

Learn the fundamentals and pros and cons of BKT and its model variants; learn how to choose between different model variants using techniques such as cross validation; learn how to apply pyBKT to real world tutoring scenarios and research questions.

Target audience: Knowledge tracing researchers, educational software developers, learning analytics professors/teachers.

1.3. pyBKT[1] Model Fitting Example

Model Fitting and Cross-validation Code:

```
>>> from pyBKT.models import Model
>>> model = Model(seed = 42, num_fits = 1)
>>> model.fit(data_path = 'cognitive_tutor.csv', skills = 'Plot pi')
>>> model.params()
```

```
                           value
skill   param   class
Plot pi prior   default 0.62830
        learns  default 0.58967
        guesses default 0.13011
        slips   default 0.09073
        forgets default 0.00000
```

```
>>> print(model.crossvalidate(data_path = 'ct.csv', skills = 'Plot pi', folds = 5))
```

```
           rmse
skill
Plot pi 0.46242
```

[1] https://github.com/CAHLR/pyBKT

REFERENCES

Pardos, Z. A., Heffernan, N. T. (2010) Modeling Individualization in a Bayesian Networks Implementation of Knowledge Tracing. In P. De Bra, A. Kobsa, D. Chin (Eds.) *Proceedings of the 18th International Conference on User Modeling, Adaptation and Personalization* (UMAP). Big Island of Hawaii. Pages. Springer. Pages 255-266.

Pardos, Z. A., Heffernan, N. T. (2010) Using HMMs and bagged decision trees to leverage rich features of user and skill from an intelligent tutoring system dataset. In J. Stamper & A. Niculescu-Mizil (Eds.) *Proceedings of the KDD Cup Workshop at the 16th ACM Conference on Knowledge Discovery and Data Mining* (SIGKDD). Washington, D.C. ACM. Pages 24-35.

Pardos, Z. & Heffernan, N. (2011) KT-IDEM: Introducing Item Difficulty to the Knowledge Tracing Model. In Konstant et al. (eds.) *Proceedings of the 20th International Conference on User Modeling, Adaptation and Personalization* (UMAP). Girona, Spain. Springer. Pages 243-254.

Pardos, Z. A., Bergner, Y., Seaton, D., Pritchard, D.E. (2013) Adapting Bayesian Knowledge Tracing to a Massive Open Online College Course in edX. In S.K. D'Mello, R.A. Calvo, & A. Olney (Eds.) *Proceedings of the 6th International Conference on Educational Data Mining* (EDM). Memphis, TN. Pages 137-144.

Yudelson, M. V., Koedinger, K. R., & Gordon, G. J. (2013, July). Individualized bayesian knowledge tracing models. In *International Conference on Artificial Intelligence in Education* (pp. 171-180). Springer, Berlin, Heidelberg.

DesignLAK21: Rapid prototyping of learning analytics visualisations for learning design

Linda Corrin
Swinburne University of Technology
lcorrin@swin.edu.au

Aneesha Bakharia
University of Queensland
a.bakharia1@uq.edu.au

Nancy Law
University of Hong Kong
nlaw@hku.hk

Sandra Milligan
University of Melbourne
s.milligan@unimelb.edu.au

Ulla Ringtved
University College of Northern Denmark
ulr@ucn.dk

ABSTRACT: The 6th Annual DesignLAK Workshop addresses a need in the learning analytics community to be able to rapidly prototype learning analytics visualisations with educators based on their needs relating to the learning designs used in their teaching context. The development of learning analytics applications for educators is often a long and complex process. It necessarily involves a range of people with skills in data science and application development and can demand multiple iterations with key stakeholders before prototypes are approved. In this interactive, 2.5 hour workshop we will explore an alternative approach to this co-design and prototyping process which aims to help educators to design learning analytics visualisations for their learning designs in a less time- and expertise-intensive way. The workshop will involve the use of a newly developed prototyping tool known as DIVE, developed at MIT, which supports data analysis and visualisation without the need for users to be able to code. Workshop participants will have a chance to use the DIVE tool to create visualisations that would be useful for their teaching contexts and to provide feedback on the process to inform further work on how this tool could be adapted to support learning analytics application development.

Keywords: Learning analytics, learning design, visualisation, prototyping

1 BACKGROUND

The development of learning analytics applications for educators is often a long and complex process. In some cases, an application is developed by an institution or vendor in a generic way so it can be used to explore data visualisations across many different learning designs (Bakharia et al., 2016). Alternatively, a more focused approach could be taken, in consultation with educators, to develop applications that are more specific to certain tools and/or learning designs within particular teaching contexts. This co-design approach has gained popularity among the learning analytics community in recent times (e.g., Dollinger et. al., 2019, Shibani et. al., 2019) and can involve several iterations of consultation with educators to develop designs for visualisation and application development. These discussions are often supported initially by the use of tools such as templates and/or cards (e.g. Alvarez et al., 2020) that help educators to think through the different design and data elements necessary for subsequent development of a prototype. Further meetings are then needed to review the translation of the design to the prototype environment, involving people with a range of data science and developer skills, before development can begin on the final application.

In this workshop we will explore an alternative approach to this co-design and prototyping process which aims to help educators to design learning analytics visualisations for their learning designs in a less time- and expertise-intensive way. The workshop will involve the use of a newly developed prototyping tool known as DIVE, developed at MIT[1], which supports data analysis and visualisation without the need for users to be able to code. This open-source tool includes many features that can help educators and learning designers to easily build useful data representations such as synthetic data generation, automated visualisation, and data model recommendations. The online workshop will involve an introduction to the development of learning analytics visualisations for learning design followed by a demonstration of the DIVE tool enhanced with extensions specifically to support the co-design of learning analytics applications. Participants will then have a chance to use the DIVE tool to create visualisations that are useful for their teaching contexts and to provide feedback on the process to inform further work on how this tool could be adapted to support learning analytics application development. This aligns with the LAK21 theme of “the impact that we make” in that our aim is to provide more efficient and cooperative ways for learning analytics visualisations to be prototyped and developed to allow educators to use data to have impact on learning and teaching practice.

The DesignLAK series of workshops have offered an opportunity over the past five years to explore the intersection of learning analytics and learning design. Earlier workshops focused on improving feedback processes (Milligan et al., 2016) and indicators for assessment design (Ringtved et al., 2017). In 2018, a series of learning analytics/learning design tools were showcased and evaluated (Corrin et al., 2018), and in 2019 the workshop considered the validity of data used for assessment analytics (Law et al., 2019). DesignLAK21 will continue this conversation around the relationship between learning analytics and learning design while augmenting this in a practical way through the development of a rapid prototype relevant to participants’ own context. The interactive nature of the workshop will provide participants exposure to a way to build learning analytics prototypes that help

[1] https://www.media.mit.edu/projects/dive/overview/

educators and learning designers to analyse and visualise data with reference to learning design in ways that can inform appropriate interventions and/or design amendments.

2 OBJECTIVE OF THE WORKSHOP

The objective of the DesignLAK21 workshop is to explore how a prototyping tool (DIVE) can be used within a co-design workshop setting to make the design of learning analytics visualisations more efficient and meaningful for/with educators and learning designers. The workshop will provide an opportunity to engage learning analytics researchers and practitioners (LAK attendees) in this process to contribute to an ongoing program of research about how co-design workshops can be enhanced with rapid prototyping to help educators to be able to align analytics with learning design to produce insights that can lead to action (Bakharia & Corrin, 2020).

3 WORKSHOP DESIGN

The DesignLAK21 workshop will take place online and will run for 2.5 hours. The event will involve a number of interactive activities including small-group collaborative tasks, as well as whole group discussions and feedback opportunities. The workshop will be facilitated by the DesignLAK team in an online synchronous tool that enables the creation of break-out groups and provides communication channels to facilitate ongoing discussions and feedback provision. The online nature of the workshop means it can be opened to slightly larger numbers than previous DesignLAK workshops (which usually average around 20 participants), however we would need to cap attendance around 40-50 participants to allow for the DesignLAK21 team to be able to facilitate and monitor the groups and provide feedback on and support for the prototype designs. The workshop will be open to any educators, learning designers, researchers, and learning analytics practitioners who have an interest in how learning analytics visualisations can be prototyped in ways that make specific reference to learning design in practice.

3.1 Pre-workshop preparation

Prior to the workshop, the organisers will promote the event through a range of social media (e.g. Twitter, Slack, etc.) and the mailing lists of associated professional societies (e.g. SoLAR, ASCILITE, etc.). The DesignLAK website will be updated to contain: a summary of the workshop design, information about the workshop facilitators, the workshop schedule, and further resources that may be of interest to participants relating to research on the visualisation of learning analytics with reference to learning design. In the lead up to the workshop an introduction email will be sent to participants to remind them of the workshop schedule and to collect information about their roles, context, and interests (this information will be used to allocate participants to groups within the break-out sessions). They will also be encouraged to consider questions they may have around their learning designs that they might like to use as the basis of a visualisation design in the workshop.

3.2 The workshop

The workshop will open with an icebreaker activity that will set the scene for an exploration of the intersection of learning design and analytics. This will then lead into a whole group discussion of important considerations for visualisation design of learning analytics data, which will be interspersed with examples from the research of how data from different learning designs can be visualised. Emphasis will be placed on the importance of storytelling techniques to enable various stakeholders (e.g. educators, learning designers, etc.) to be able to interpret these visualisations so that actions can be planned and enacted. Participants will then be introduced to the DIVE tool and the main functionality will be demonstrated. They will then be divided into small groups of 5-6 participants to work on generating a visualisation design using DIVE around a learning design that would be relevant to their own practice. Synthetic datasets will be provided for use within the DIVE tool and group members will be given the option to either work on their own designs, or to work together as a group on a design of common interest (where possible, participants with similar interests will be grouped together on the basis of the information they provided prior to the workshop). A DesignLAK organiser will be allocated to facilitate each break-out group room, but may need to move between groups if the participant numbers are large. A 15 min break will be scheduled during the design section of the workshop to give participants a chance to rest, eat, and/or move around - or they could opt to use the extra time to work on their designs. At the end of the design section, the whole group will come back together to share designs and to receive feedback on these designs from other participants. The workshop will end with a feedback generation activity about the process of the prototype development as well as opportunities and challenges that they may have identified through taking part in this process.

4 WORKSHOP OUTCOMES

The workshop offers benefits to both the participants and organisers. Participants will gain experience in using the DIVE tool to preview recommended visualisations, and to assemble and share selected visualisations as data stories related to their learning designs. The organisers will be able to consider the visualisation artefacts, participant discussions, and feedback to evaluate the use of DIVE as a low-code rapid prototyping and co-design tool. Throughout the workshop participants will be encouraged to Tweet about their experience, using the hashtag #DesignLAK21, to disseminate awareness of the session and any important ideas and/or provocations that emerge from the discussion. A summary of the session resources and resulting discussions will be compiled by the DesignLAK21 organisers and made available on the website post the event. Permission will be sought from willing participants to include examples of resulting visualisations on the website. The formal evaluation of the workshop will form part of an ongoing study of this rapid prototyping approach for learning analytics being conducted by members of the DesignLAK21 organising group and so this workshop will feature in resulting publications (appropriate ethical approval for the use of workshop data will be obtained prior to the workshop and communicated with participants). As usual, ideas of interest that emerge from the workshop will be used to inspire the focus of DesignLAK22.

REFERENCES

Alvarez, C. P., Martinez-Maldonado, R., & Shum, S. B. (2020, March). LA-DECK: A card-based learning analytics co-design tool. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 63-72). https://doi-org/10.1145/3375462.3375476

Bakharia, A., Corrin, L., de Barba, P., Kennedy, G., Gasevic, D., Mulder, R., Williams, D., Dawson, S., Lockyer, L. (2016). A conceptual framework linking learning design with learning analytics. In T. Reiners, B.R. von Konsky, D. Gibson, V. Chang, L. Irving, & K. Clarke (Eds.), Proceedings of the 6th International Conference on Learning Analytics and Knowledge (pp. 409-413). New York: ACM. http://dx.doi.org/10.1145/2883851.2883944

Bakharia, A. & Corrin, L. (2020). Reimagining the conceptualisation, design and delivery of Learning Analytics applications. In So, H. J. et al. (Eds.), *Proceedings of the 28th International Conference on Computers in Education* (pp. 215-224). Asia-Pacific Society for Computers in Education.

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.

Dollinger, M., Liu, D., Arthars, N., & Lodge, J. (2019). Working Together in Learning Analytics Towards the Co-Creation of Value. *Journal of Learning Analytics, 6*(2), 10–26. https://doi.org/10.18608/jla.2019.62.2

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., 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.

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.

Shibani, A., Knight, S., & Shum, S. B. (2020). Educator perspectives on learning analytics in classroom practice. *The Internet and Higher Education*, *46*, 100730. https://doi.org/10.1016/j.iheduc.2020.100730