# LAK24 Conference Proceedings

*Learning Analytics in the Age of Artificial Intelligence*

# The Fourteenth International Conference on Learning Analytics & Knowledge

March 18-22, 2024,
Kyoto, Japan

**Organized by:** SOLAR Society for Learning Analytics Research

# LAK24 Conference Proceedings

*Learning Analytics in the Age of Artificial Intelligence*

The Association for Computing Machinery
1601 Broadway, 10th Floor
New York, New York 10019, USA



**ACM ISBN:** 979-8-4007-1618-8

# LAK24 Conference Proceedings

*Learning Analytics in the Age of Artificial Intelligence*

## LAK24 Program Chairs' Welcome

We are very pleased to welcome you to the Fourteenth International Conference on Learning Analytics and Knowledge (LAK24), organized by the Society for Learning Analytics Research (SoLAR). With the aim of widening participation of the Learning Analytics (LA) community, this year's conference is held between March 20th and 22th.

The theme for the 14th annual LAK conference is ***Learning Analytics in the Age of Artificial Intelligence***. Artificial intelligence has been relevant for learning analytics since the early days of the field. This has mostly been manifested by building upon the algorithms of machine learning to analyze data about learners and learning environments. The conversations about artificial intelligence in education used to be mostly contained within specialized communities of practitioners and researchers. Since late 2022, this rapidly changed. Discourse in mainstream media and among the general public has been dominated by the coverage of the developments in generative artificial intelligence. The notable examples are technologies such as ChatGPT and DALL-E that harness the power of deep learning algorithms to generate impressively human-like text and images based on relatively simple human prompts. These technologies have given some glimpses about the emerging age of artificial intelligence and the profound impact it will have on research and practice in education.

Three excellent keynotes address various aspects of artificial intelligence in education and its relationship to learning analytics. Mutlu Cukurova is a Professor of Learning and Artificial Intelligence at University College London (UCL). Mutlu's keynote delves into the interplay of AI and LA looking at the potentials, pitfalls and the future of education. Dr. Kristen DiCerbo, Chief Learning Officer at Khan Academy, brings a practitioner perspective and gives insights into implementing AI in a learning environment at scale. Her keynote focuses on using AI to enhance human intelligence, informing us on how to use dialogic interaction to better understand students' thought processes and helping educators better understand what learners know and can do. Professor Stephen J.H. Yang is the Vice President for R&D and Chair Professor of Computer Science and Information Engineering at National Central University, Taiwan. Stephen's keynote explores the potential of generative AI and LLMs in learning analytics, highlighting the challenges that need to be addressed and the opportunities they bring to enhance educational outcomes. There will also be a panel highlighting efforts on the practical implementation of learning analytics at scale in Japanese schools. The panel, *Connecting Research, Practice and Policies for Large-scale Learning Analytics*, comprises researchers, policy makers, educational technology companies and teachers and students from a K12 school in Kyoto. The conference features other panels that are focused on the conference theme (learning analytics in the age of AI) and strengthening collaboration links across different stakeholder groups in learning analytics.

This year's conference theme encouraged researchers and practitioners to consider the implications for learning analytics and the role the field can play in the age of artificial intelligence. We received a very large number of high-quality submissions this year breaking all previous records, and we are extremely grateful for all those who have contributed to our LAK24. The research track had 316 submissions (205 full paper submissions and 111 short paper submissions). This represents an increase of about 41% in the total number of submissions compared to last year. Maintaining the high quality of the conference, the program committee for the research track consisted of 287 researchers from the field of learning analytics, educational data mining, learning sciences, educational technology, and related disciplines. Of these, 47 are senior members, all recognized leaders in the field and highly involved in service to the learning analytics community. Overall, from the 316 research submissions, the program committee worked very hard to select 95 papers (66 full research papers and 29 short research papers) that are included in these proceedings of the Learning Analytics and Knowledge Conference. The overall acceptance rate for the conference was 30%, while the acceptance rates for the full and short research tracks were 32.7% and 26.1%, respectively.

The rigorous selection process for LAK includes an initial phase of review of at least two program committee members. Authors are then given a short time to provide an optional rebuttal to the remarks and comments raised in the initial review in which they can answer specific questions raised by reviewers (if any) or provide clarifications and justifications. Each submission is then carefully reviewed by one of the senior program committee members who provides a summary meta-review and final recommendation to the program chairs, based on both original reviews and any rebuttal submitted by authors. We are most grateful for all the hard work by the program committee and their insightful and constructive comments and reviews. These proceedings could not have been possible without their generous help and support.

We would also like to emphasize our ongoing gratitude for the efforts made by all involved in our community. The past few years have been difficult due to the ongoing impact of COVID. We very much understand the complexity of work and life pressures impacting on our time commitments, and priorities. The high level of support and commitment shown by our colleagues to ensure that the presented and published papers have received high quality reviews and feedback is highly valued and appreciated. These are difficult times for us all and we want to thank you for the important efforts you have devoted that have allowed this conference to continue as a premier scientific event fostering the scholarly exchange of ideas of the highest caliber.

We hope that LAK24 participants and other readers of these proceedings will find value in the many varied contributions to the field of LA contained within. The prominence of artificial intelligence has also opened profound debates about implications on education from the need to develop relevant literacies to work with artificial intelligence, to challenging the established notions of assessment in education. We hope that this conference encourages researchers to consider implications for learning analytics and the role the field can play in the age of artificial intelligence.

**Brendan Flanagan**
*Kyoto University, Japan*

**Barbara Wasson**
*University of Bergen, Norway*

**Dragan Gašević**
*Monash University, Australia*

# LAK24 Conference Proceedings

*Learning Analytics in the Age of Artificial Intelligence*

## LAK24 Conference Organization

**Program Chairs**
Brendan Flanagan, Kyoto University, Japan
Barbara Wasson, University of Bergen, Norway
Dragan Gašević, Monash University, Australia

**Practitioner Track Chairs**
Elizabeth Dalton, Intelliboard, United States
Danny Liu, University of Sydney, Australia

**Workshop Chairs**
Dilrukshi Gamage, Tokyo Institute of Technology, Japan
Gloria Fernández Nieto, Monash University, Australia
Atsushi Shimada, Kyushu University, Japan

**Posters and Demonstrations Chairs**
Xiao Hu, University of Hong Kong, Hong Kong
Ramkumar Rajendran, Indian Institute of Technology Bombay, India

**Doctoral Consortium Chairs**
Christopher Brooks, University of Michigan, USA
Jelena Jovanović, University of Belgrade, Serbia
Negin Mirriahi, University of South Australia, Australia
Luis Prieto, Universidad de Valladolid, Spain

**Proceedings Chairs**
Srecko Joksimovic, University of South Australia, Australia
Andrew Zamecnik, University of South Australia, Australia

**Website Manager**
Yancy Vance Paredes, Arizona State University, USA

**Organizing Committee**
Grace Lynch, Society for Learning Analytics Research, Australia
Hiroaki Ogata, Kyoto University, Japan
Brendan Flanagan, Kyoto University, Japan
Nicole Hoover, Society for Learning Analytics Research, USA

## LAK24 Program Committee

### Senior Program Committee

Roger Azevedo, University of Central Florida, United States
Ryan Baker, University of Pennsylvania, United States
Gautam Biswas, Vanderbilt University, United States
Christopher Brooks, University of Michigan, United States
Simon Buckingham Shum, University of Technology, Sydney, Australia
Cassandra Colvin, University of Tasmania, Australia
Tinne De Laet, Katholieke Universiteit Leuven, Belgium
Yannis Dimitriadis, University of Valladolid, Spain
Hendrik Drachsler, DIPF, Germany
Shihui Feng, The University of Hong Kong, Hong Kong
Rebecca Ferguson, The Open University, United Kingdom
Rafael Ferreira Mello, Universidade Federal Rural de Pernambuco, Brazil
Michail Giannakos, Norwegian University of Science and Technology, Norway
Geraldine Gray, Technological University Dublin, Ireland
Marek Hatala, Simon Fraser University, Canada
Davinia Hernandez-Leo, Universitat Pompeu Fabra, Barcelona, Spain
Pedro J. Muñoz-Merino, Universidad Carlos III de Madrid, Spain
Srecko Joksimovic, Education Future, University of South Australia, Australia
Jelena Jovanovic, University of Belgrade, Serbia
Sanna Järvelä, University of Oulu, Finland
Mohammad Khalil, University of Bergen, Norway
Hassan Khosravi, The University of Queensland, Australia
Kirsty Kitto, University of Technology, Sydney, Australia
Simon Knight, University of Technology, Sydney, Australia
Kenneth Koedinger, Carnegie Mellon University, United States
Vitomir Kovanovic, The University of South Australia, Australia
James Lester, North Carolina State University, United States
Alejandra Martinez Mones, University of Valladolid, Spain
Roberto Martinez-Maldonado, Monash University, Australia
Manolis Mavrikis, The University of Edinburgh, United Kingdom
Agathe Merceron, Berliner Hochschule für Technik - University of Applied Sciences, Germany
Negin Mirriahi, University of South Australia, Australia
Inge Molenaar, Radboud University, Netherlands
Xavier Ochoa, New York University, United States
Abelardo Pardo, University of Adelaide, Australia
Mykola Pechenizkiy, Eindhoven University of Technology, Netherlands
Oleksandra Poquet, Technical University of Munich, Germany
Paul Prinsloo, University of South Africa, South Africa
Bart Rienties, The Open University, United Kingdom
Maren Scheffel, Ruhr University Bochum, Germany
John Stamper, Carnegie Mellon University, United States
Yi-Shan Tsai, Monash University, Australia
H. Ulrich Hoppe, University Duisburg-Essen / RIAS Institute Duisburg, Germany / Kyoto University, Japan
Katrien Verbert, Katholieke Universiteit Leuven, Belgium
Olga Viberg, KTH Royal Institute of Technology, Sweden
Phil Winne, Simon Fraser University, Canada
Alyssa Wise, Vanderbilt University, United States

# LAK24 Conference Proceedings

*Learning Analytics in the Age of Artificial Intelligence*

**Program Committee**

José A. Ruipérez Valiente, University of Murcia, Spain
Stephen Aguilar, University of Southern California, Rossier School of Education, United States
Samantha Ahern, University College London, United Kingdom
Nora'Ayu Ahmad Uzir, Universiti Teknologi MARA, Malaysia
Laia Albó, Universitat Pompeu Fabra, Spain
Sarah Alcock, The Open University, United Kingdom
Giora Alexandron, Weizmann Institute of Science, Israel
Riordan Alfredo, Monash University, Australia
Laura Allen, University of Minnesota, United States
Dana Alzoubi, Mississippi State University, United States
Ishari Amarasinghe, Universitat Pompeu Fabra, Spain
Mohammad Amin Samadi, University of California, Irvine, United States
Mohamed Amine Chatti, University of Duisburg-Essen, Germany
Máverick André Dionísio Ferreira, Federal University of Pernambuco, Brazil
Kimberly Arnold, University of Wisconsin-Madison, United States
Thushari Atapattu, The University of Adelaide, Australia
Renato Azevedo, University of Illinois Urbana-Champaign, United States
Aneesha Bakharia, The University of Queensland, Australia
Arun Balajiee Lekshmi Narayanan, University of Pittsburgh, United States
Tiffany Barnes, North Carolina State University, United States
Carla Barreiros, Know-Center GmbH, Austria
Abhinava Barthakur, University of South Australia, Australia
Alan Berg, University of Amsterdam, Netherlands
Maurice Boothe, New York University, United States
Conrad Borchers, Carnegie Mellon University, United States
Nigel Bosch, University of Illinois Urbana-Champaign, United States
François Bouchet, Sorbonne Université - LIP6, France
Alex Bowers, Columbia University, United States
Andrew Brackett, New York University, United States
Julien Broisin, Université Toulouse 3 Paul Sabatier - IRIT, France
Olga C. Santos, UNED, Spain
Dan Carpenter, North Carolina State University, United States
Cristian Cechinel, Universidade Federal de Santa Catarina, Brazil
Teresa Cerratto-Pargman, Stockholm University, Sweden
Li Chen, Faculty of Arts and Science, Kyushu University, Japan
Guanliang Chen, Monash University, Australia
Weiqin Chen, Oslo Metropolitan University, Norway
Lujie Chen, University of Maryland, Baltimore County, United States
Yixin Cheng, Monash University, Australia
Marco Chiarandini, University of Southern Denmark, Denmark
Heeryung Choi, Massachusetts Institute of Technology, United States
Irene-Angelica Chounta, University of Duisburg-Essen, Germany
Linda Corrin, Deakin University, Australia
Yi Cui, Capilano University, Canada
Mutlu Cukurova, University College London, United Kingdom
Cynthia D'Angelo, University of Illinois at Urbana-Champaign, United States
Yiling Dai, Kyoto University, Japan
Crina Damsa, University of Oslo, Norway
Daniel Davis, Delft University of Technology, Netherlands
Lars de Vreugd, University Medical Centre Utrecht, Netherlands
Carrie Demmans Epp, University of Alberta, Canada
Professor Denise Whitelock, The Open University, United Kingdom
Daniele Di Mitri, DIPF | Leibniz Institute for Research and Information in Education, Germany
Amber Dood, University of Michigan, United States

Mohsen Dorodchi, University of North Carolina Charlotte, United States
Brendan Eagan, University of Wisconsin-Madison, United States
Martin Ebner, Graz University of Technology, Austria
Vanessa Echeverria, Monash University, Australia
Anthony F. Botelho, University of Florida, United States
Yizhou Fan, The University of Edinburgh, United Kingdom
Stephen Fancsali, Carnegie Learning, Inc., United States
Yael Feldman-Maggor, Weizmann Institute of Science, Israel
Gloria Fernández Nieto, Monash University, Australia
Nivan Ferreira, Universidade Federal de Pernambuco, Brazil
Albrecht Fortenbacher, HTW Berlin, Germany
Stephanie Frost, University of Saskatchewan, Canada
Paula Galvao de Barba, Monash University, Australia
Bhagya Gangodagamachchige, Monash University, Australia
Kerr Gardiner, Jisc, United Kingdom
Xuewang Geng, Graduate School of Human-Environment Studies, Kyushu University, Japan
Carlos Gerardo Prieto Alvarez, The University of Sydney, Australia
Andrew Gibson, Queensland University of Technology, Australia
Sebastian Gombert, DIPF | Leibniz Institute for Research and Information in Education, Germany
Sabine Graf, Athabasca University, Canada
Dai Griffiths, UNIR-iTED, Spain
Maureen Guarcello, San Diego State University, United States
Julio Guerra, Universidad Austral de Chile, Chile
Ashish Gurung, Carnegie Mellon University, United States
Eduardo Gómez-Sánchez, Universidad de Valladolid, Spain
Elaine H. T. Oliveira, Universidade Federal do Amazonas, Brazil
Ella Haig, School of Computing, University of Portsmouth, United Kingdom
Eyad Hakami, Pompeu Fabra University, Spain
Scott Harrison, German Institute for International Educational Research, Germany
Carl Haynes-Magyar, Carnegie Mellon University, United States
Neil Heffernan, Worcester Polytechnic Institute, United States
Birte Heinemann, RWTH Aachen University, Germany
Christothea Herodotou, The Open University, United Kingdom
Isabel Hilliger, Pontificia Universidad Católica de Chile, Chile
Martin Hlosta, The Swiss Distance University of Applied Sciences, Switzerland
Caitlin Holman, University of Michigan, United States
Izumi Horikoshi, Kyoto University, Japan
Sharon Hsiao, Santa Clara University, United States
Xiao Hu, The University of Hong Kong, Hong Kong
Bowen Hui, The University of British Columbia, Canada
Paul Hur, University of Illinois at Urbana-Champaign, United States
Stephen Hutt, University of Denver, United States
Juan I. Asensio-Pérez, Collaborative and Intelligent Systems Group, Universidad de Valladolid, Spain
Sehrish Iqbal, Monash University, Australia
Yoon Jeon Kim, University of Wisconsin-Madison, United States
María Jesús Rodríguez-Triana, Tallinn University, Estonia
Ioana Jivet, Goethe University Frankfurt & DIPF, Germany
Song Ju, North Carolina State University, United States
Yeonji Jung, University of Memphis, United States
Rogers Kaliisa, University of Oslo, Norway
Onur Karademir, DIPF | Leibniz Institute for Research and Information in Education, Germany
Shamya Karumbaiah, University of Wisconsin-Madison, United States
Rene Kizilcec, Cornell University, United States
Justian Knobbout, HU University of Applied Sciences, Netherlands
Elizabeth Koh, National Institute of Education, Nanyang Technological University, Singapore, Singapore

Malgorzata Korolkiewicz, UniSA, Australia
Huan Kuang, Florida State University, United States
Jakub Kužílek, CIIRC, CTU in Prague, Czechia
Tanja Käser, EPFL, Switzerland
Miguel L. Bote-Lorenzo, Universidad de Valladolid, Spain
Sébastien Lallé, Sorbonne University, France
Andrew Lan, University of Massachusetts Amherst, United States
Charles Lang, Columbia University, United States
Eitel Lauria, Marist College, United States
Elise Lavoué, iaelyon, Université Jean Moulin Lyon 3, LIRIS, France
Armanda Lewis, New York University, United States
Tobias Ley, University for Continuing Education Krems, Austria
Lin Li, Monash University, Australia
Tongguang Li, Monash University, Australia
Xinyu Li, Monash University, Australia
Yuheng Li, Monash University, Australia
Ziyi Li, Monash University, Australia
Huiyong Li, Research Institute for Information Technology, Kyushu University, Japan, Japan
Zhi Li, University of California, Berkeley, United States
Chenglu Li, University of Florida, United States
Fanjie Li, Vanderbilt University, United States
Changhao Liang, Kyoto University, Japan
Lyn Lim, Technical University of Munich, Germany
Lisa Lim, University of Technology Sydney, Australia
Jionghao Lin, Carnegie Mellon University, United States
Yiwen Lin, University of California, Irvine, United States
Qinyi Liu, University of Bergen, Norway
Hyunbin Loh, Riiid, South Korea
Vanda Luengo, Sorbonne Université - LIP6, France
Joni Lämsä, University of Oulu, Finland
Sonsoles López-Pernas, University of Eastern Finland, Finland
Leah Macfadyen, The University of British Columbia, Canada
Rwitajit Majumdar, Kumamoto University, Japan
Jorge Maldonado-Mahauad, Universidad de Cuenca, Pontificia Universidad Católica de Chile, Chile
Juan Manuel Dodero, Universidad de Cádiz, Spain
Pedro Manuel Moreno-Marcos, Universidad Carlos III de Madrid, Spain
Mirko Marras, University of Cagliari, Italy
Wannisa Matcha, Prince of Songkla University, Thailand
Diego Matos, Federal University of Alagoas, Brazil
Rebecca Matz, University of Michigan, United States
George Michailidis, University of Florida, United States
Konstantinos Michos, University of Zurich, Switzerland
Caitlin Mills, University of Minnesota, Canada
Tsubasa Minematsu, Kyushu University, Japan
Kamila Misiejuk, University of Bergen, Norway
Steven Moore, Carnegie Mellon University, United States
Hanni Muukkonen, University of Oulu, Finland
Alison Myers, The University of British Columbia, Canada
Andre Nascimento, UFRPE, Brazil
Debarshi Nath, IIT Bombay, India
Sadia Nawaz, Monash University, Australia
Tanya Nazaretsky, EPFL, Israel
Huy Nguyen, Carnegie Mellon University, United States
Ha Nguyen, Utah State University, United States
Patrick Ocheja, Kyoto University, Japan

Fumiya Okubo, Kyushu University, Japan
Alejandro Ortega-Arranz, Universidad de Valladolid, Spain
Zacharoula Papamitsiou, SINTEF Digital, Norway
Luc Paquette, University of Illinois at Urbana-Champaign, United States
Zach Pardos, University of California, Berkeley, United States
Yeonjeong Park, Honam University, South Korea
Ward Peeters, Kanda University of International Studies, Japan
Debbie Perouli, Marquette University, United States
Michelle Perry, University of Illinois at Urbana-Champaign, United States
Malte Persike, RWTH Aachen University, Germany
Andrew Petersen, University of Toronto, Canada
Taciana Pontual Falcão, Universidade Federal Rural de Pernambuco, Brazil
Stanislav Pozdniakov, Monash University, Australia
Sambit Praharaj, Center for Advanced Internet Studies and Ruhr University Bochum, Germany
Ying Que, The University of Hong Kong, China
Joshua Quick, Indiana University Bloomington, United States
Sabina Rako, University Computing Centre University of Zagreb, Croatia
Mladen Rakovic, Monash University, Australia
Cheng Ren, University of California, Berkeley, United States
Fernando Rodriguez, University of California, Irvine, United States
Ido Roll, Technion - Israel Institute of Technology, Israel
Daniela Rotelli, Università di Pisa, Italy
Jonathan Rowe, North Carolina State University, United States
Merike Saar, Tallinn University, Estonia
Shazia Sadiq, The University of Queensland, Australia
John Saint, Regents University, United Kingdom
Majd Sakr, Carnegie Mellon University, United States
Fatemeh Salehian Kia, University of British Columbia, Canada
Jeanette Samuelsen, University of Bergen, Norway
Agnes Sandor, Naver Labs Europe, France
Mohammed Saqr, University of Eastern Finland, Finland
Marcel Schmitz, Zuyd Hogeschool, Netherlands
Clara Schumacher, Humboldt Universität zu Berlin, Germany
Niels Seidel, FernUniversität in Hagen, Germany
Lele Sha, Monash University, Australia
Xuequn Shang, Northwestern Polytechnical University, China
Kshitij Sharma, Norwegian University of Science and Technology, Norway
Yang Shi, North Carolina State University, United States
Antonette Shibani, University of Technology, Sydney, Australia
Atsushi Shimada, Kyushu University, Japan
Hong Shui, University of Minnesota, United States
Diana Simic, University of Zagreb, Faculty of Organization and Informatics, Croatia
Shaveen Singh, University of the South Pacific, Fiji
Sergey Sosnovsky, Utrecht University, Netherlands
Daniel Spikol, University of Copenhagen, Denmark
Namrata Srivastava, Monash University, Australia
Elisabeth Staudegger, Universität Graz, Austria
Zhiru Sun, University of Southern Denmark, Denmark
Zachari Swiecki, Monash University, United States
Kyosuke Takami, Education Data Science Center, National Institute for Educational Policy Research (NIER), Japan
Yuta Taniguchi, Kyushu University, Japan
Michelle Taub, University of Central Florida, United States
Mohammadreza Tavakoli, TIB, Germany
Kp Thai, Age of Learning, United States

Craig Thompson, The University of British Columbia, Canada
Jeremy Tzi-Dong Ng, The University of Hong Kong, Hong Kong
Anouschka van Leeuwen, Utrecht University, Netherlands
Joep van der Graaf, Radboud University, Netherlands
Lorenzo Vigentini, The University of New South Wales, Australia
Ignacio Villagran, Pontificia Universidad Católica de Chile, Chile
Dai Wei, Monash University, Australia
Jacob Whitehill, Worcester Polytechnic Institute, United States
John Whitmer, Blackboard, Inc., United States
Kimberly Williamson, Cornell University, United States
Yang Xu, The Policy Lab at Brown University, United States
Masanori Yamada, Kyushu University, Japan
Lixiang Yan, Monash University, Australia
Iman Yeckehzaare, University of Michigan, United States
Vwen Yen Alwyn Lee, Nanyang Technological University, Singapore
Renzhe Yu, Teachers College, Columbia University, United States
Zhang Yupei, Northwestern Polytechnical University, China
Kateryna Zabolotna, University of Oulu, Finland
Angela Zavaleta Bernuy, University of Toronto, Canada
Linxuan Zhao, Monash University, Australia
Gaoxia Zhu, National Institute of Education (NIE), Nanyang Technological University, Singapore
Xinran Zhu, University of Pennsylvania, United States

***Practitioner Reports***

# Patterns of Learning Behaviors 4 Success in Asynchronous Professional Training 4 Educators

**Dean Cristol, PhD**
The Ohio State University
cristol.2@osu.edu

**Belinda Gimbert, PhD**
The Ohio State University
gimbert.1@osu.edu

**ABSTRACT**: The growing prevalence of online learning has resulted in an increased reliance on big data and learning analytics, which generate digital traces capturing students' engagement, performance, and learning experiences. Utilizing this learning analytics data enables the extraction of features and the construction of models to understand the learning process, particularly focusing on measuring students' engagement through quantitative measures of behavioral engagement. This presentation explores various learning patterns observed among participants in a professional institute engaged in self-directed online learning and discusses how these patterns influence the mastery of knowledge and skills. Drawing inspiration from Wigfield & Eccles' (2000) expectancy-value theory model, the presentation connects the motivational theoretical principles of the model to comprehend the relationship between motivation and engagement. Employing data visualization techniques, the presenters have identified five distinct learning patterns within the online context and have pinpointed various learning strategies employed by the participants.
**Keywords:** Student Learning and Teaching Processes, Educational Policies and Strategies, Digital Traces, Learning Patterns, Expectancy-Value Theory

## 1 CONTEXT

The expansion of online education prompts an upswing in data collection and the application of learning analytics to monitor students' digital interactions, performance, and progress (Sin & Muthu, 2015). Behavioral engagement indicators, such as frequency and time-on-task, are pivotal in gauging students' interaction with learning materials (Nakamura et al., 2021). Higher frequency signifies more active engagement, and time-on-task measures the duration spent on specific learning activities, with longer durations reflecting increased dedication to learning tasks (Wong & Chong, 2018). This growth in online education facilitates data-driven insights into student engagement, with researchers prioritizing behavioral indicators for analyzing digital interactions and time allocation to learning activities (Vytasek et al., 2020). The use of learning analytics by extracts valuable insights into engagement, performance, and learning processes, enabling data-driven interventions, and refining educational practices for more effective and personalized learning experiences (Piotrkowicz et al., 2021). The guiding research question examines how different learning patterns emerge as participants engage with the required learning activities of the Professional Development Institute (PDI), an asynchronous training initiative tailored for individuals seeking a teaching license and designed to furnish vital knowledge and skills crucial for proficient classroom instruction. We explored the intricate connections between these learning patterns, the delivery of content, and the mastery of knowledge and skills in the context of self-directed learning.

# 2 BACKGROUND

Continuous learning empowers educators to adeptly navigate and thrive amidst the constantly evolving demands of education, guaranteeing their efficacy in nurturing student development and achievement (Darling-Hammond & Richardson, 2009). The learner-centered approach recognizes the distinctive learning requirements of professionals and emphasizes that they are most capable of identifying and meeting these needs. Leveraging learner-centered approaches and technology enables professionals to assume command of their learning journey, adjusting and progressing in response to the dynamic demands of their respective fields (Archambault, 2022).

Expectancy-Value Theory a Model in an Online Learning Context: The expectancy-value theory model proposed by Wigfield & Eccles (2000) offers a robust framework for comprehending motivation and engagement in educational settings. This theoretical framework revolves around two fundamental components: expectancy beliefs and value beliefs. Expectancy beliefs pertain to students' perceptions of their capability to excel in various learning activities. In the realm of online learning, students hold specific expectations regarding their technological proficiency, access to course materials, and their ability to complete assignments, all of which significantly influence their motivation to actively engage with the learning process (Niven, 2022). Value beliefs encapsulate the significance that students attribute to tasks or learning activities. In the realm of online learning, students evaluate the value of learning materials, the relevance of content to their objectives, and the utility of online platforms (Edwards & Taasoobshirazi, 2022). The application of the expectancy-value theory to online learning yields valuable insights into the motivational factors that impact student engagement and achievement. This understanding serves as a guide for crafting effective online learning experiences, and empowering educators and instructional designers to construct supportive and engaging online environments that elevate student motivation and enhance learning outcomes.

# 3 RESULTS

## 3.1 The Frequency Patterns of Behavioral Engagement

The dataset comprised 2803 task sessions, where assignments and overview tasks predominated, and information tasks were less frequent. Session durations ranged from 1 to 160 clicks, with an average of 9 clicks and a median of 5 clicks, indicating brief sessions. Noteworthy is that 412 sessions had only 1 click, while the remaining 2391 sessions involved multiple clicks. After removing duplicates, 1485 unique sessions remained. This finding suggests that despite 87 participants contributing 2391 sessions with two or more actions, there were only 485 distinct behavioral patterns exhibited.

## 3.2 Hidden Markov Model (HMM) (clusters; overall model descriptives)

Applying Markov Chains, we treated the sessions generated by students as a stochastic process, opting for a 1st-order Markov chain where the state at time n depends solely on the recent state. This choice was deemed optimal based on the matrices of AIC and BIC. For each session, we created 1st-order transition matrices, capturing the probabilities of transitioning between states, taking into account the current state (Si) and the previous state (Si-1). With 5 pages in consideration, a 5x5 transition matrix was generated for each session, representing the session as a 25-dimensional vector. To cluster these sessions, the k-means algorithm was employed, iterating through each session, and calculating the Euclidean distance between the session's transition matrix row and each cluster centroid. Subsequently, the session was assigned to the cluster with the closest centroid. The mean of the transition matrix rows within each cluster was calculated, and these steps were repeated until cluster assignments remained unchanged or reached a maximum number of iterations. The elbow method was utilized to determine the optimal number of clusters for session clustering by plotting the number of clusters against the sum of within-cluster distances. Sessions with a length of over 2 were clustered into 5 groups, labeled as Cluster 1 to Cluster 5. Additionally, the 4 types of single page click behaviors were included as separate clusters, resulting in a total of 9 clusters identified at the session level.

### 3.3 Five Unique Sequential Profiles of Behavioral Engagement

**Cluster 1,** the most widespread among all clusters, comprises 461 unique sessions, accounting for 20.7% of the total sessions. It predominantly centers on assignments and is marked by behaviors such as viewing assignment pages (34%), overview pages (29%), and grade report pages (24%). These sessions exhibit the lengthiest average duration, reaching 12.5 clicks. The assignment-to-reading ratio stands at 2.87, indicating a notable emphasis on assignment-related activities. Notably, there is a robust bidirectional connection between assignment and grade report pages, suggesting a propensity to review feedback while actively engaging with assignments.

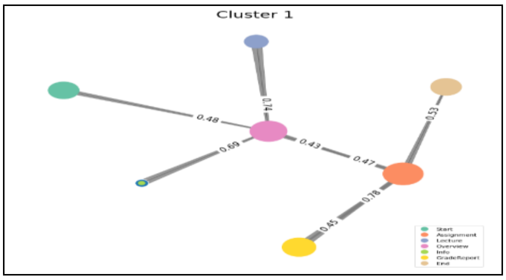



**Figure 1: Example of a network diagram for Cluster 1**

**Cluster 2** consists of 291 unique sessions, representing 12% of the sessions & presents a more balanced pattern compared to Cluster 1. These sessions exhibit an average length of 9.4 clicks. Click behaviors are characterized by 46% of views on overview pages, 33% on assignment pages & 19% on lecture pages. The assignment-to-reading ratio is 1.73, indicating a slight inclination towards assignment-related activities over reading. Robust bidirectional connections are observed between overview and assignment pages, and between lecture and overview pages. In contrast to Cluster 1, Cluster 2 features fewer clicks on the grade report page, suggesting that participants did not extensively review grades or feedback but rather focused on interactions with lecture, overview, and assignment pages.

**Cluster 3** comprises 187 unique sessions, representing 19% of all identified sessions, with the shortest average session length at 4.5 clicks. Similar to Clusters 1 and 2, it demonstrates an assignment-focused pattern, where views of assignment and overview pages constitute 84% of the total click behavior, while lecture pages account for only 14%. The assignment-to-reading ratio is 2.97, indicating a notable emphasis on assignment-related activities over reading. A robust bidirectional connection is observed between overview and assignment pages. Given the shorter session length, there is an almost 50% probability that a session either starts with the overview page or ends with the assignment page.

**Cluster 4** includes 139 unique sessions, making up 5% of all identified sessions, with an average session length of 11.6 clicks, indicating longer durations. This cluster displays a balanced pattern between assignment and lecture activities, where assignment page views constitute 25% of the total click behavior, and lecture pages account for 17%. The assignment-to-lecture ratio is 1.47, signifying a higher emphasis on assignment-related activities. Sessions within this cluster typically initiate with the overview page and then navigate to other course pages.

**Cluster 5,** the most prevalent cluster type, consists of 403 unique sessions, making up 32% of the total sessions. With an average length of 7.4, indicating a moderate duration, this cluster is characterized by an assignment-focused pattern with elements of reflection. Assignment and grade report page views collectively account for 65% of the total click behavior, while lecture pages represent 16%. The assignment-to-lecture ratio is 1.79, highlighting an emphasis on assignment-related activities. A robust bidirectional connection is observed between grade report and assignment pages, suggesting that participants within this cluster focused intently on assignments and grades.

# 4 CONCLUSION

We explored the impact of participants' learning patterns within a self-directed online learning environment on their mastery of knowledge and skills over a year-long training period. Employing Markov Chains, the analyses revealed five distinct clusters. Cluster 1 demonstrated participants engaging with feedback messages while completing assignments, often concluding sessions with the assignment page. Cluster 2 depicted a different approach, with participants focusing less on reading grades or feedback and more on switching between lecture, overview, and assignment pages. Cluster 3 indicated an equal likelihood of sessions starting with the overview or ending with the assignment. Cluster 4 sessions initiated with the overview page before exploring other content, while Cluster 5 highlighted a strong emphasis on assignments and grades. These findings underscore the diverse learning experiences within self-paced online environments. Participants' self-regulated learning skills were found to enhance their performance, aligning with Wigfield & Eccles (2000) expectancy-value theory. Participants exhibited various behaviors and successfully completed the training. By applying expectancy-value theory to online learning, motivational factors and behavioral patterns contributing to participant engagement and success were identified, providing insights for designing effective online learning experiences and environments. This study, however, has limitations such as a small dataset that could be a foundation for future studies with larger samples. The inclusion of only successful participants may introduce bias, necessitating future research to compare the learning behaviors of completers and non-completers. While learning patterns may not directly impact performance, they may correlate with motivation and self-regulation ability, prompting the collection of self-reporting data through a follow-up survey in future research.

# 5 REFERENCES

Archambault, L., Leary, H., & Rice, K. (2022). Pillars of online pedagogy: A framework for teaching in online learning environments. *Educational Psychologist, 57*(3), 178-191.

Darling-Hammond, L., & Richardson, N. (2009). Research review/teacher learning: What matters. *Educational leadership, 66*(5), 46-53.

Edwards, O. V., & Taasoobshirazi, G. (2022). Social presence and teacher involvement: The link with expectancy, task value, and engagement. *The Internet and Higher Education, 55*, 100869.

Kovanovic, V., Gašević, D., Dawson, S., Joksimovic, S., & Baker, R. (2015). Does time-on-task estimation matter? Implications on validity of learning analytics findings. *Journal of Learning Analytics, 2*(3), 81-110.

Nakamura, S., Phung, L., & Reinders, H. (2021). The effect of learner choice on L2 task engagement. Studies. Second Language Acquisition, 43(2), 428-441.

Niven, M. R. (2022). Academic procrastination and achievement motivation: An expectancy value approach to measure academic procrastination in university online courses (Doctoral dissertation, Queen's University (Canada).

Piotrkowicz, A., Wang, K., Hallam, J., & Dimitrova, V. (2021). Data-driven exploration of engagement with workplace-based assessment in the clinical skills domain. *International Journal of Artificial Intelligence in Education, 31*, 1022-1052.

Sin, K., & Muthu, L. (2015). Application of big data in education data mining and learning analytics—a literature review. *ICTACT Journal on Soft Computing, 5*(4), 1035-1049.

Vytasek, J. M., Patzak, A., & Winne, P. H. (2020). Analytics for student engagement. In *Machine Learning Paradigms: Advances in Learning Analytics* (pp. 23-48).

Wong, A., & Chong, S. (2018). Modelling adult learners' online engagement behavior: Proxy measures and its application. *Journal of Computers in Education, 5*, 463-479.

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. Contemporary *Educational Psychology, 25*(1), 68-81.

# Co-Designing Augmented Reality Interventions to Support Executive Function for Neurodiverse STEM Students

**Zachary Alstad**
TERC Inc.
zac_alstad@terc.edu

**Ibrahim Dahlstrom-Hakki**
TERC Inc.
ibrahim_dahlstrom-hakki@terc.edu

**ABSTRACT**: This practitioner report focuses on the use of a co-design approach to develop a data-driven augmented reality tool for understanding and shaping how neurodivergent students engage with their STEM-related homework. Learning analytics that are available in augmented reality headsets are robust, including head position data, head orientation data, and eye and hand tracking information. With this rich information available, there is promise in smart interventions for this population to implement more robust learning and engagement analytics. Augmented reality offers a new means of providing these interventions to students in highly customizable ways. Moreover, this sensor suite can be used to train a supervised machine-learning model to predict off-task behavior and intervene before a student is disengaged. Taken in its entirety, this system can be understood as a "smart pomodoro timer" and shows promise for this population.

**Keywords**: Training dataset; Neurodiversity; Co-design; Augmented Reality; Machine Learning; Executive Function

## 1 INTRODUCTION

There is a precedent for research exploring the utility of prompting for students with executive function (EF) challenges (Hutt et al., 2021). Automated prompting-related interventions with this population are particularly challenging due to the complexities of how they initiate and maintain engagement with the material. This complexity offers the opportunity for new and more innovative interventions. If there is a tool that can effectively detect off-task behavior, then there may be ways to more effectively intervene.

Part of the complexity of this space involves promoting these students to re-engage while avoiding simultaneously inducing frustration. Students with EF issues are more prone to distraction (Anselm et al., 2021), daydreaming (Theodor-Katz et al., 2022), inappropriate social behavior in the classroom, and behavioral disciplinary consequences such as detention or expulsion (Brandt et al., 2019). While each of these issues has its own complex social and cultural context, effective interventions at the initial point of disengagement may prevent more severe consequences for the student further down the line.

### 1.1 Augmented reality tools

An augmented reality (AR) tool can be understood as a device that takes information from a user's immediate environment and physical state, and adds informational layers on top of what they are experiencing. One example would be navigation information overlaid for a driver in unfamiliar territory. One form factor that has grown increasingly pervasive is the AR headset that one can wear on their head like a pair of glasses. These systems may process and record information related to a user's head position, head orientation, eye tracking, and hand tracking information. Each of these behavioral markers has a complex relationship to cognition and what can be understood about that individual's engagement. Because of this robust, real-time data processing, there is opportunity in this space to detect off-task behavior and potentially intervene for neurodiverse students.

### 1.2 Neurodiversity

For the purposes of this work, neurodiversity refers primarily to those individuals who struggle with EF barriers, a source of challenge for many individuals including (but not limited to) those with autism, ADHD, and learning disabilities. EF barriers include a range of challenges that can significantly impact learning. These may include difficulty initiating work, maintaining focus, shifting focus, self-regulation, managing frustration, and difficulties with memory (Brown, 2006). These areas of challenge are shared by a broad range of individuals who may struggle in learning environments due to underlying neurological differences, psychological trauma, or brain trauma. The neurodiversity movement espouses the ethos, “Nothing about us without us” (Charlton, 1998). Here, we report on our experience co-designing a learning analytics-based intervention with neurodivergent students.

## 2 CO-DESIGN

A number of design approaches utilize feedback from their intended audience to guide their work. This can range from traditional data collection, simple surveys, focus groups, and participatory design all the way to full co-design. Co-design as we define it here incorporates the target audience in all aspects of the project from the outset. All aspects of this project are guided by and incorporate the perspectives and participation of the neurodivergent community. This includes having neurodivergent members on the research team, as well as working closely with neurodivergent students when designing the learning analytics, informed studies, and interventions.

### 2.1 Training Dataset

To train the ML component of the intervention, participants were recruited at a campus that exclusively serves neurodivergent learners and asked to do their homework as normal while wearing AR headsets as they were being recorded. All participants volunteered to participate and received a gift card for their time in accordance with an IRB approved protocol. Off-task behavior was then manually coded, as is evidenced in the video data. This was understood as any non-task related behavior that lasted more than one second. Each of these human-coded, off-task events was validated by a second coder. Data collection is ongoing and focuses on developing and refining an ML algorithm to detect off-task behavior.

### 2.2 Off-task detection algorithm

Development of the off-task detection algorithm will involve the use of transformer models trained using the dataset described in section 2.1. Co-design has supported this aspect of development as well. The project team has explored what level of temporal granularity to use for these models. In addition, co-designers have helped develop the coding scheme for on- and off-task behavior and whether to code the behavior using a binary coding scheme or one that includes additional states (e.g., unsure, somewhat distracted, etc.).

### 2.3 Interface & Prompt Design

A significant aspect of the co-design process focuses on the development of an effective AR interface that can prompt users appropriately. While the detection of off-task behavior is challenging, one could argue that responding appropriately and effectively is of greater importance. The biggest potential liability for such an AR system is that it may negatively impact the performance of its users if its prompts interrupt periods of productive behavior or cause frustration to the user. That is why the prompts designed must be false-alarm averse and appropriate to the learner's needs at that moment.

Co-designers have helped us consider a range of non-intrusive prompts to subtly alert the user to changes in attention without explicitly interrupting their work. For example, they have suggested the inclusion of background white noise or music to block out distracting sounds. Changes in the type, tempo, or volume of that background noise/music can be used to alert the user to changes in their on-task behavior. Another suggestion included dimming or occluding parts of the visual field when a user is thought to be off task to help reduce visual distractions in the periphery and redirect their attention.

#### Lessons Learned

The co-design process is critical to this project's ability to serve its intended audience effectively and appropriately. However, developers and designers who want to employ co-design must dedicate the necessary resources to ensure that it is a productive endeavor for both the project and the co-designers. In that spirit, we share the following lessons learned:

- **Co-designer goals may not fully overlap with project goals**: Many co-designers are interested in developing soft skills, exploring career paths, or simply learning more about an interesting topic area
- **Co-design takes time**: Growing your development teams means more voices need to be considered and many co-designers are new to this process
- **Co-designers and team members bring different skills/knowledge to the project**: There will be a range of personal experiences and expertise on the team that must be bridged

## 3 Recommendations

While this work is still ongoing, there are some actionable recommendations that have manifested from these efforts. With the rise of technologies that capture attention, subvert executive function, and influence participation in often coercive ways, it is evident there is a need for more tools that allow people to take control of their own agency. Those with executive function issues are particularly sensitive to this sort of influence. Initial work with this population and anecdotal feedback from participants suggests that prompting of this kind in an AR environment would be beneficial when students are struggling to initiate and maintain focus.

Inclusion of co-designers in this process has yielded important insights into the dynamics of engagement and distraction. This process has been fruitful in related work and the process has been articulated in greater detail (Dahlstrom-Hakki et al., 2021). Co-designers have offered insights into the practicality (or lack thereof) of AR and the hardware itself. One important insight was related to audio information in that the inclusion or omission of audio such as music can have wide individual differences with regards to preference. Some students need ambient audio to remain engaged. In order to re-engage these students, removing the ambient audio can call attention to their behavior. Conversely, other students will need silence to remain engaged and, should they break attention away from their target, an interruption of an audio cue is more appropriate.

Based on the aforementioned project goals, we recommend the following:

- Ensure that the project provides opportunities for co-designers to explore career paths, develop soft skills, and pursue their own academic and career goals through the project.
- Build in the time and resources to build new connections across your team, explore the most effective means of communications, and build a trusting working relationship.
- Find ways to ensure that team members bring their expertise to the table. It will take time for co-designers to feel empowered to help shape the process. It can be difficult to ensure that more experienced team members do not dominate the conversation but still contribute enough of their expertise to help the process move forward effectively.

## REFERENCES


Anselm B. M. F., Tucha, L., Butzbach, M., Weisbrod M., Aschenbrenner, S., & Tucha, O. (2021). ADHD at the workplace: ADHD symptoms, diagnostic status, and work-related functioning. *J Neural Transm* (Vienna). 128(7): 1021–1031. https://doi: 10.1007/s00702-021-02309-z.

Brandt, V., Koerner, J. K., & Palmer-Cooper, E. (2019). The Association of Non-obscene Socially Inappropriate Behavior With Attention-Deficit/Hyperactivity Disorder Symptoms, Conduct Problems, and Risky Decision Making in a Large Sample of Adolescents. *Front Psychiatry*; 10: 660.

Brown, T. E. (2006). *Attention deficit disorder: The unfocused mind in children and adults*. Yale University Press.

Charlton, J. I. (1998). Nothing about us without us. University of California Press. ISBN 0-520-22481-7.

Dahlstrom-Hakki, I., Edwards, T., Larsen, J., Alstad, Z., Belton, G., Lougen, D., & Santana, D. (2021). Inclusive VR through inclusive co-design with neurodiverse learners. 7th International Conference of the Immersive Learning Research Network (17 May 2021–10 June 2021). Eureka, CA, USA.

Hutt, S., Krasich, K., Brockmole, J. R., & D'Mello, S. K. (2021). Breaking out of the lab: Mitigating mind wandering with gaze-based attention-aware technology in classrooms. CHI Conference on Human Factors in Computing Systems, Yokohama, Japan. https://doi.org/10.1145/3411764.3445269.

Theodor-Katz, N., Somer, E., Hesseg, R. M., Soffer-Dudek, N. (2022). Could immersive daydreaming underlie a deficit in attention? The prevalence and characteristics of maladaptive daydreaming in individuals with attention-deficit/hyperactivity disorder. *J Clin Psychol*; 78(11): 2309–2328. https://doi:10.1002/jclp.23355

# A Roadmap for Implementing Learning Analytics: taking into account Pedagogy, Privacy, Ethics, and Technical Infrastructure

**Anouschka van Leeuwen**
Utrecht University
a.vanleeuwen@uu.nl

**Marije Goudriaan**
Utrecht University
m.goudriaan1@uu.nl

**Ünal Aksu**
Utrecht University
u.aksu@uu.nl

**Momena Yousufzai**
Utrecht University
m.yousufzai1@uu.nl

**ABSTRACT**: This practitioner report details the development of a roadmap for initiating and evaluating Learning Analytics (LA) projects that were created at Utrecht University. The aim for the roadmap was to enable a combination of a top-down and bottom-up approach to LA: by top-down providing facilities with governance steps, the goal was to enable educators at the University to bottom-up initiate LA projects. The report starts with a description of the context of Utrecht University and the need to develop the roadmap. Subsequently, we discuss the groundwork that was required at the institutional level, for example, establishing an institutional policy for LA. Then, we present the roadmap and our experience with its application so far. The roadmap may be of use to other institutions as an example of how to combine a top-down and bottom-up approach to LA.

**Keywords**: Roadmap, Institutional policy, Learning analytics, Educational data

## 1 CONTEXT AND MOTIVATION FOR A LEARNING ANALYTICS ROADMAP

In an increasingly data-driven world, educational institutions are challenged to position themselves in how they deal with educational data and for what purposes they allow such data to be used. Many educational institutions recognize that data offer the opportunity to gain insights into and improve processes concerning teaching and learning, and have initiated efforts to implement learning analytics (LA) (Gasevic et al., 2019). Implementing LA is a complex endeavor that requires developments in the areas of pedagogy, ethics, privacy, and technology (Gasevic et al., 2019; Lonn et al., 2017; Ifenthaler et al., 2021), and a continuous consideration of stakeholder input (Gray et al., 2022).

In this practitioner report, we discuss the development of LA in our University. Two years ago, a central LA team was composed to develop the facilities needed to implement LA in a pedagogically and ethically sound way. The goal was to combine a top-down and bottom-up approach (Perez-Sanagustin et al., 2022). The top-down approach consisted of ensuring that all needed facilities were in place at the central level, such as an LA policy and an LA technical infrastructure. The bottom-up approach consisted of controlling the initiation of new LA projects at the staff level within faculties, thereby ensuring stakeholder engagement. To bridge the central and faculty levels and enable faculties to realize new LA projects, a *Roadmap for LA* (from now on called the Roadmap) was created that outlines the steps needed. Thus, staff can follow these steps and obtain support from the central level for an LA project. The remainder of this paper elaborates on the Roadmap (section 2) and discusses recommendations based on our experience with the Roadmap (Section 3). Thereby, we provide the Roadmap as inspiration and basis for other educational institutions.

## 2 THE ROADMAP

### 2.1 Preparations at the Institutional Level

As indicated, the choice was made to combine a top-down and bottom-up approach to LA. At the institutional (i.e., central) level, facilities were created, empowering the initiation and implementation of LA within the faculties. On the technical side, this included the development of a data platform suitable for storage, processing, analyzing, and visualizing educational data for various projects. On the pedagogical, ethical, and legal side, there was a need to create a central LA policy. We will highlight the LA policy here because it had direct consequences for the Roadmap.

The LA policy was inspired by a national reference framework[1], and existing literature on LA policies at other educational institutions (Tsai & Gasevic, 2017). This was combined with internal conversations with the following stakeholders: teachers, students, program managers, vice deans, and privacy officers. The policy was finalized and approved by the University board and consisted of three parts. The first part outlined pedagogical goals for LA, indicating for which purposes LA could be applied. Four application areas were outlined: (1) course-level real-time LA, (2) LA aimed at course evaluation and curriculum development, (3) LA for competence-based education, and (4) LA for predicting study delay and success. In the second part of the policy, five key ethical values were established: transparency, responsibility, fair consideration of stakeholder interests, reliable and valid analyses, and keeping humans in the loop. In the third part, the legal framework for LA was outlined, including adherence to European privacy regulations (GDPR) and the need to perform a privacy security check on every LA project. The policy can be found at our website[2].

### 2.2 Presenting the Roadmap

The Roadmap is visualized in Figure 1. Based on feedback from teachers on an initial version, it was presented in five clear steps, with an accompanying website[2] that lists the sub-steps with supporting materials (e.g., templates). The Roadmap indicates in broad terms the goal of each step and provides an estimation of the time that each step will take. The five steps are explained in more detail below.

[1] https://www.versnellingsplan.nl/Kennisbank/referentiekader-privacy-en-ethiek-voor-studiedata/

[2] https://www.uu.nl/en/education/learning-analytics/la-in-education

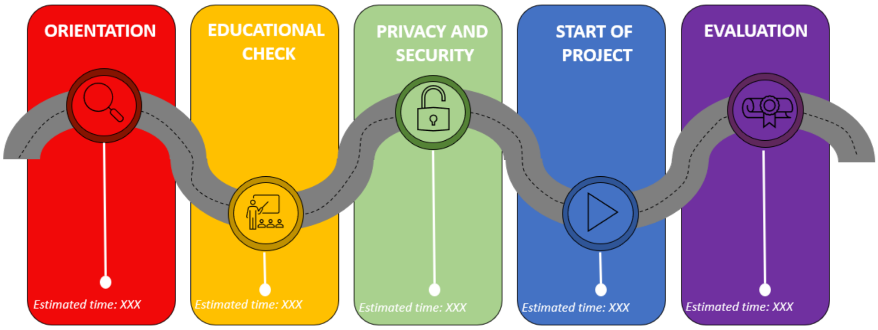



Figure 1: Roadmap for LA Projects

**(1) Orientation**: the project initiators, usually faculty staff, seek contact with the central LA team to propose their project. The team then requests information about the project's aims and hypotheses, potential data sources, the desired data analyses, and how the results should be presented to end-users (e.g., via a dashboard). With this information, the central team makes an informed decision on whether the project is feasible.

**(2) Educational check:** the project is pitched to an educational committee consisting of several stakeholder representatives (student, teacher, policy maker, researcher, innovation manager). The committee checks the proposed project from several perspectives: why it is relevant, whether it adheres to the pedagogical goals of the central LA policy, and how it will be evaluated. Afterward, the committee will decide if the project can start or what adjustments are needed.

**(3) Privacy and security:** a privacy scan is carried out to determine the project's risk level regarding data and privacy regulations by weighing all stakeholder perspectives. A data protection impact assessment (DPIA) needs to be carried out in case of a high risk. The scan is followed by a technical risk assessment to determine the security strength of the project's dataflows and tools. This technical risk level dictates the steps that need to be taken to ensure that security is in line with the regulations in place, such as the university's privacy regulations and the GDPR.

**(4) Start of project:** once the above-explained checks are performed, the data owner is asked for consent to start the project. At this point, the actual technological developments can start, such as a student-facing dashboard. Further, for all LA projects, the project initiator will inform all related project stakeholders before the start of the project using a privacy statement.

**(5) Evaluation:** the goal of the central LA team is to provide support to LA projects, and one way to do so is by evaluating projects and learning from them. Accordingly, project initiators are asked to provide an evaluation plan for the educational check step. Once the project has officially started, collecting the measures for evaluation can begin. The evaluation outcomes will serve as a basis for the decision whether to embed the developed LA project in the University's education and expand it to a broader audience with a long-term use purpose. Moreover, the evaluation outcomes are essential to determine the required changes if the project does not deliver the expected results.

## 3 EXPERIENCES WITH THE ROADMAP

At the time of writing, the Roadmap has been in place for six months. Overall, our experience is positive because the Roadmap offers clarity and ensures that all stakeholders' perspectives are heard and weighed. An important new insight we gained is related to data governance. Using LA at our University required an adaptation of the current IT infrastructure. For example, the current infrastructure has allowed for storing structured data, while in LA projects, unstructured data is often used. In our case, this led to the development of a LA-specific data platform flexible for data storage, processing, analysis, and visualization possibilities. The new data platform also led to questions concerning data governance in terms of who maintains access to data and who is responsible for data processing. Since we did not include this aspect in our central policy, our recommendation would be to answer these questions in the central LA policy or an overarching data governance document for the whole institution.

The primary benefit of the Roadmap has been that all required steps, roles and responsibilities are transparent, and the staff proposing a new LA project knows what to expect. It also shows the complexity and the array of expertise needed for LA, which means staff requesting a project may require additional resources. Therefore, in the orientation step, we always recommend project leaders to apply for innovation funding (either within our University or externally) to ensure they have dedicated time to engage in an LA project. Hopefully, with more experience, the five steps of the Roadmap become more efficient because we can build on example cases. To sum up, based on our experience, we list the following recommendations:

1. A LA Roadmap helps to identify and balance perspectives of all involved stakeholders
2. A LA Roadmap helps to build clarity by communicating policy, roles and responsibilities
3. Technology and agreements on data governance are an important pillar of data-driven innovation
4. Dedicated time for LA project leaders is essential to carry out all steps, including evaluation

## REFERENCES


Gasevic, D., Tsai, Y. S., Dawson, S., & Pardo, A. (2019). How do we start? An approach to learning analytics adoption in higher education. *International Journal of Information and Learning Technology*, *36*(4), 342–353. https://doi.org/10.1108/IJILT-02-2019-0024

Gray, G., Cooke, G., Murnion, P., Rooney, P., & O'Rourke, K. C. (2022). Stakeholders' insights on learning analytics: Perspectives of students and staff. *Computers & Education*, *187*(May), 104550. https://doi.org/10.1016/j.compedu.2022.104550

Ifenthaler, D., Gibson, D., Prasse, D., Shimada, A., & Yamada, M. (2021). Putting learning back into learning analytics: actions for policy makers, researchers, and practitioners. *Educational Technology Research and Development*, *69*(4), 2131–2150. https://doi.org/10.1007/s11423-020-09909-8

Lonn, S., McKay, T. A., & Teasley, S. D. (2017). Cultivating institutional capacities for Learning Analytics. *New Directions for Higher Education*, *2017*(179), 53–63. https://doi.org/10.1002/he.20243

Perez-Sanagustin, M., Hilliger, I., Maldonado-Mahauad, J., & Perez-Alvarez, R. (2022). Building institutional capacity for Learning Analytics: Top-down & bottom-up initiatives. *IEEE*, 1–9. https://doi.org/10.1109/RITA.2022.3191413

Tsai, Y., & Gasevic, D. (2017). Learning Analytics in Higher Education – challenges and policies : a review of eight Learning Analytics policies. *LAK '17: Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, 233–242. https://doi.org/10.1145/3027385.3027400

# Co-LA: A Course-Learner Analytics framework for Co-instructing with Generative Language Models in Computing Education

[1]**Lele Sha,** [1]**Zhiping Liang,** [1]**Yuheng Li,** [1]**Kaixun Yang,** [1]**Linxuan Zhao,** [1]**Lixiang Yan,** [1]**Gloria Fernandez Nieto,** [1]**Zachari Swiecki,** [1]**Shuai Fu,** [1]**Guanliang Chen,** [1]**Dragan Gašević,** [1]**Xinyu Li**
[1]Monash university
[lele.sha,zhiping.liang,yuheng.li,kaixun.yang1,linxuan.zhao,lixiang.yan,gloriamilena.fernandeznieto,zach.swiecki,shuai.fu,guanliang.chen,dragan.gasevic,xinyu.li1]@monash.edu

**ABSTRACT**: Instructions are essential to guide novice students to navigate complex concepts in computing education. While recent advancements in generative large language models (LLMs) have made it possible to scale up the creation of instructional content. However, it is important to acknowledge that a general-purpose LLM comes with intricacies and context-dependent nuances that require recognition and careful responses in educational settings. To this end, we propose Co-LA, a learning analytics-based framework aimed at facilitating the co-instruction process between teachers and LLM that recognizes the course context while responding to learners' fine-grained knowledge gaps. We illustrate the analytics of course, learner, and various interaction modes that the Co-LA framework affords in an authentic university course and show how these can facilitate and reveal constraints on LLM's co-instruction capabilities in education.

**Keywords**: Computing Education, Large Language Model, Instruction

## 1 INTRODUCTION

In recent studies, consensus seems to have emerged regarding the potential of generative LLMs (e.g., Codex, GPT-3, etc.) to automate the production of readily usable instructions in computing education (e.g., programming exercises, code explanations, exemplars) (Finnie-Ansley et al., 2022; Sarsa et al., 2022). Notably, these studies have two limitations. 1) ***Inadequate assessment of LLM's capability in educational settings***, usually pivoting around LLM and not on the learner (e.g., what the latest LLM can do that previous AI cannot). But how informative is such assessment for educators? 2) ***Too much emphasis on writing the right prompt***, and too little on the co-creation process, which could distill valuable co-instruction insights for the teaching team. These motivated the Co-LA framework – a Course-Learner Analytics framework that situate instruction co-creation within the course context to address learner's instructional needs. Co-LA is distinguished from prior work in the following respects. First, Co-LA is a *human-centered Learning Analytics framework*, aiming to equip educators to conduct effective co-instruction with LLM, rather than automating the instruction creation. Second, Co-LA effectively situates LLM co-instruction within fine-grained course-learner context, enabling the measurement of learning gains facilitated by the instruction process, to better inform educational stakeholders. Finally, Co-LA does not dictate on the correct use of prompts, instead, focusing on Human-AI interaction process that can elicit the optimal co-instruction to produce effective and scalable instructional content.

## 2 CO-LA FRAMEWORK

Co-LA is composed of three inter-dependent processes: *Course analytics*, *Learner analytics*, and *Co-instruction*. The goal is to situate the co-instruction process within the course context and tailored to the learner. First, *Course analytics* set the course instruction context on a weekly basis via descriptive analytics of different instructional materials. Operationally, it is most efficient to derive all course instructional materials in a systematic top-down manner, e.g., a programming topic may have many concepts, which may be instructed using quiz questions and / or programming exercises. This way, the course instruction could be represented using meaningful textual data to both teacher and LLMs. *Learner analytics* ensures that instruction disseminated to a learner can address the learner's instructional needs, characterized by the triple (cognitive load, knowledge gap, learning pattern), which are all identified using Learning Analytics techniques. This helps to formulate instruction that are at the 1) right timing; 2) with the right content; and 3) with the right instructional strategy. Finally, *Co-instruction* is a process where the analytics results are passed to teacher and LLM, creating a "*shared mental model*" of the course instruction context and learner's needs between both parties, so they are informed of 1) whether a new instruction needs to be created (e.g., new knowledge gap identified); and 2) existing instruction needs to be revisited (e.g., students are confused about a piece of code). We illustrate the operationalization of Co-LA in the next section.

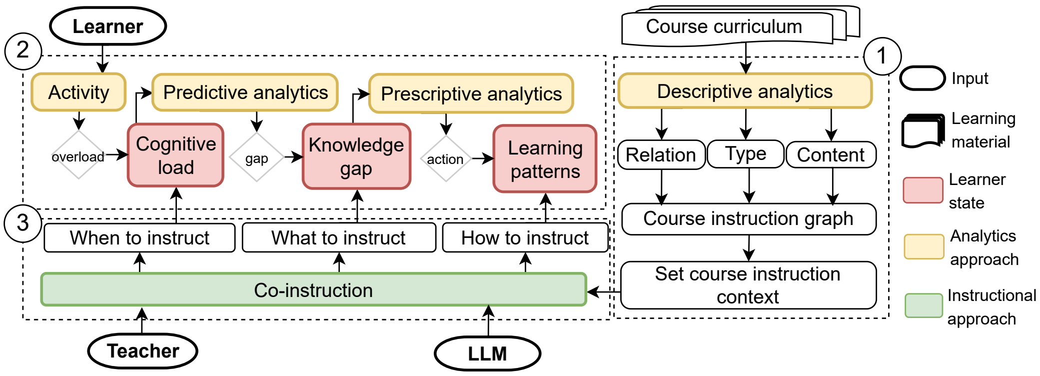



**Figure 1: Illustrating the Co-LA framework.**

## 3 RESULTS

We applied Co-LA to a large-scale offering (500 students) of the introductory Python programming course at the Monash university. The course is carried out in a weekly manner (12 weeks per offering), and students are expected to participate in an in-person workshop session and pre/post-workshop online learning activities on Ed Discussion (an LMS commonly used in STEM education).

**Course analytics**. We worked closely with teaching team and systematically extracted the instructional content in pre-class, in-class and post-class activities, which can be conceptually represented as an entity-relation structure (denoted to as *course instruction graph*) and easily stored/retrieved in relational database (see Figure 2). For example, a programming concept belongs to one curriculum topic and may have 0 or many coding exercises. The course instruction graph is dynamic and continuously updated each week as the course progresses. For any instructional needs, existing course instructional materials could be easily retrieved from the database along with its related topic / concept as course instructional context, which could either be prompted as an example case for LLM on a newly detected knowledge gap (e.g., confusion on exception handling), or be revised for different learning needs (e.g., more detailed line-by-line code explanation).

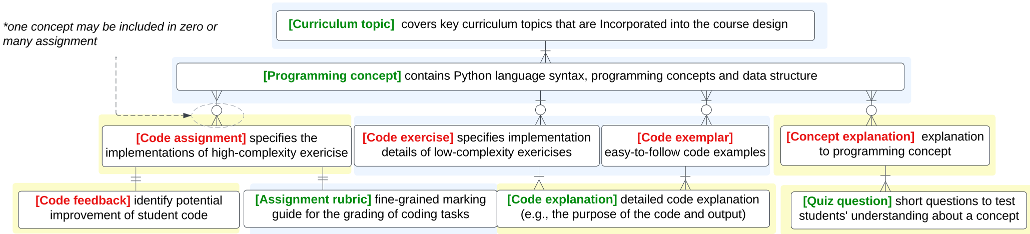



**Figure 2: Summative view of course instruction graph. The relation is based on [instruction type]. green: low replacement rate; red: high replacement rate; yellow: assessable; blue: non-assessable.**

**<u>Learner analytics</u>**. To characterize student instructional needs, 35 features (Fincham et al., 2018) were engineered from student activity trace on Ed Discussion, e.g., time spent on coding tasks and pre/post-workshop material. To assess cognitive load, we identified students who spent high effort (time and content access) while achieving low performance in coding exercises (determined by a senior teaching staff). To identify knowledge gaps, we employed a supervised predictive modelling approach on student assessment data and code, using a pretrained CodeBERT model (Feng et al., 2020) to automatically classify various misconceptions about taught programming knowledge (about 71% accuracy). We found that most students (with medium-high cognitive load) displayed 2-5 distinct knowledge gaps each week. Finally, a hidden markov model (HMM) was adopted to encode low-level learner activity trace into high-level learning patterns, to inform teachers about learner's instructional needs, e.g., if an active student (i.e., completing assigned tasks on Ed Discussion) still displayed code logical error, all teachers agreed that the best strategy was to break down the coding block into different logical components with detailed reasoning process to scaffold problem-solving, a few representative examples are summarized in Table 1.

**Table 1: Learner analytics and corresponding instructional strategy.**

| Load | Gap | Patterns | Instructional strategy |
|---|---|---|---|
| High | Coding logic | Active on exercises | Scaffold coding logic via reasoning and exemplars |
| Medium | Variable scope | Active on lectures | Present coding exercises with sample solutions |
| Low | Constructor | Minimally active | Revisit taught concepts with detailed explanations |

**<u>Co-creation observations</u>**. For top-50 Co-LA identified distinct (load, gap, pattern) triple and the corresponding course instructions, we conducted offline co-creation sessions between 5 teachers and ChatGPT. First, while prior research in computing education largely concluded that LLM generations may be readily usable in practice (Finnie-Ansley et al., 2022; Sarsa et al., 2022), we observed following failure cases where the LLM output cannot be informative and/or may be harmful to serve as a basis for co-instruction, despite significant progress being made, i.e., from codex to GPT4: 1) ***Topic drifting*** when creating high-complexity coding exercises that included multiple implementations of coding constructs, the output could be drifting to undesirable / unexpected topic; 2) ***Incorrect / vague explanations*** on certain concepts and coding construct that contradicted with course teaching; 3) ***Wrong grain-size*** feedback / reasonings on codes were too high-level and not informative. This highlighted the importance of assessing LLM capability in addressing ***authentic learner knowledge***

***gaps*** (enabled by Co-LA) rather than generating generic instructional content (e.g., coding explanations that do not address learner's confusion). Given that the quality of LLM generation has been developing rapidly, we advocate for assessing LLM capability not in terms of how "novel" the generation is, but how much it can be situated within the curriculum and tailored to the learner. Second, for the successful co-creations, although teachers generally agreed that it could be helpful to students, they were less certain about whether such content can effectively produce learning for students compared to traditional instruction. This highlighted the importance to ***explore student learning process*** under hybrid human-AI instruction settings. So future studies may exploit Co-LA to assess comparative learning gains (i.e., filling knowledge gap) under traditional instruction vs. human-AI co-instruction. Finally, we note that Co-LA does NOT dictate on ***mode of interaction*** between Human-AI, as we observed (usually) a mix of teacher-initiated, collaborative, and LLM-initiated (LLM prompts teacher to write instruction) modes within a single co-creation session, see Figure 3. Across all modes, Co-LA act as a *shared mental model* that facilitated a shared understanding about a specific learner-instruction scenario, subsequently, new instructional ideas emerge (e.g., apply pseudo code to scaffold student problem-solving). This implies that co-instruction with LLM may be a multifaceted process, analogous yet distinct from human-human collaboration – there still remains complexities to be unpacked, particularly the co-ideation process, which may be distilled into a set of co-instruction skills via Co-LA's process of *when, what and how to instruct learner* to be shared among teaching team.

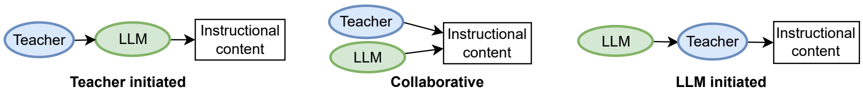



**Figure 3: Three modes of interaction**

## 4 CONCLUSION

Fundamentally, the Co-LA framework illuminate how to enhance learning with learning analytics in a post-LLM era, where instruction creation may be scaled and personalized at a low cost. By involving teacher in the process of co-instruction, we hope to regulate this process – optimizing learning gain for students while minimizing any potential harm. Ultimately, through Co-LA, we envision this novel LLM technology being successfully embedded within formal education to benefit all stakeholders.

## REFERENCES

Feng, Z., Guo, D., Tang, D., Duan, N., Feng, X., Gong, M., ... & Zhou, M. (2020). Codebert: A pre-trained model for programming and natural languages. *arXiv preprint arXiv:2002.08155*.

Fincham, E., Gašević, D., Jovanović, J., & Pardo, A. (2018). From study tactics to learning strategies: An analytical method for extracting interpretable representations. *IEEE Transactions on Learning Technologies*, *12*(1), 59-72.

Finnie-Ansley, J., Denny, P., Becker, B. A., Luxton-Reilly, A., & Prather, J. (2022, February). The robots are coming: Exploring the implications of openai codex on introductory programming. In *Proceedings of the 24th Australasian Computing Education Conference* (pp. 10-19).

Sarsa, S., Denny, P., Hellas, A., & Leinonen, J. (2022, August). Automatic generation of programming exercises and code explanations using large language models. In *Proceedings of the 2022 ACM Conference on International Computing Education Research-Volume 1* (pp. 27-43).

# Enhancing Workplace Training Efficiency: An Exploration of Employees' Behavioral Patterns on Training Outcome Using Learning Analytics

**WeiHan Lin[1], Xingheng Wang[1], Xiaoshan Huang[2], Mengyu Liang[1], Yihua Wu[3]**
[1.]Novartis Innovative Medicine, China; [2.] McGill University, Canada; 3. Shanghai Delightgo Internet Technology Co,. Ltd., China
harrylin@umich.edu

**ABSTRACT**: The extensive adoption of online learning has yielded significant volumes of learning data, offering organizations an opportunity to validate the effectiveness of online learning enhancements. However, it remains uncertain how employees can derive greater benefits from online training with increased efficiency. To address this issue, we conducted a study involving 1791 employees from a large multinational pharmaceutical company in professional skills online training over a fourteen-month period. We analyzed the trace data, employed cluster analysis to evaluate metrics related to behavioral patterns, learning outcomes, and business performance. We identify three distinct groups in terms of patterns of learning regularity, and training outcomes. Among these three groups, employees who exhibited greater regularity in their practice times also achieved higher monthly learning scores and better business performance within the workplace. Furthermore, employees with higher regularity in online training also achieve higher learning efficiency during the training program. The benefit of this study is two-folded: first, our study demonstrates the effectiveness of utilizing learning analytics in understanding employees' behavioral patterns within workplace training contexts; second, the findings have important practical implications for promoting regularity in online professional skills development programs, thereby enhancing workplace training efficiency and contributing to business performance.

**Keywords**: Clustering analysis, Learning regularity, Behavioral pattern, Training efficiency, Business performance.

## 1 INTRODUCTION

In the pursuit of improving workplace training efficiency, the objective is for employees to achieve significant professional development outcomes and elevate business performance with minimal time investment in practice. This endeavor raises the crucial question of how to optimize online training efficiency when it is entirely self-paced (Wang et al., 2022). This study employs trace data and learning analytics to delve into the intricacies of employees' learning behaviors within a learning management system. Our aim is to uncover patterns and behaviors that are indicative of enhanced learning and business performance. We apply a machine learning technique and employ principal component analysis (PCA), followed by k-means clustering to sort out key behavioral features in workplace learning. Through this rigorous analytical approach, we discern three distinct groups of employees with their self-paced regularity patterns of learning and training outcome. Furthermore, analysis of variance (ANOVA) reveals that individuals in the group with higher regularity in learning also achieve significantly higher learning and business performance than the other group(s).

## 2 METHOD

### 2.1 Participants and Procedure

The study involved 1,791 individuals employed at a large multinational pharmaceutical company in China. For inclusion in the analysis, we considered individuals who possessed complete performance and clock-in records spanning twelve or more months, from March 2022 to May 2023. The company aims to elevate the professional knowledge of medical representatives dealing with constant medical updates through a learning management system (LMS) designed for knowledge reinforcement program encompassing Disease Knowledge, Product Knowledge, and Market Competitors Knowledge. Medical representatives have access to the learning management system for online training between 7:00 AM and 11:00 PM, facing ten customized questions based on their learning records. After each correct answer, the question will reappear after a gap of 4 days for the second time. If answered correctly again, it will appear for the third time after a 2-day gap. To enhance learning efficiency, if a question is answered correctly on three consecutive attempts, that question enters a dormant state. The program strikes a balance between flexibility and consistency, mandating 80% engagement on working days. Representatives complete around 600 questions, tailored to their products, including reinforcements, within fourteen months.

## 2.2 Data Source, Measures, and Analysis

Our data sources are extracted from the LMS logfiles. It includes time series features (e.g., time of learning on each day), consisting of 200 to 300 daily logs of each employee. Training outcome is indicated by learning (outcome and efficiency) and business performance. Learning outcome is indicated by the average test score of learning by month (Mean score) while learning efficiency is indicated by the count of dormant questions (count_dorm). The higher the count of dormant questions, the greater the learning efficiency achieved. Business performance is indicated by Percentile Rank within the regional manager (RM) completion rate (RM_rank_per); the smaller number indicates a higher rank in performance. To analyze our data, we employ a multi-step approach (Figure 1). We apply feature engineering, a machine learning technique to identify implicit time-series features in employees' behavioral patterns. Using 'tsfresh' python package (Chris et al., 2018), we extract large-scale time series features and sort out relative important features to learning. We then use PCA, a dimensionality reduction technique, to filter and streamline our data in determining the optimal number of clusters for subsequent analysis (Abdi and Williams, 2010). Following PCA, we proceed to cluster our data using selected variables. This step allows us to group individuals with similar characteristics, revealing underlying patterns within our dataset. Finally, we conduct ANOVA to quantify the significance of the differences observed across these clusters.

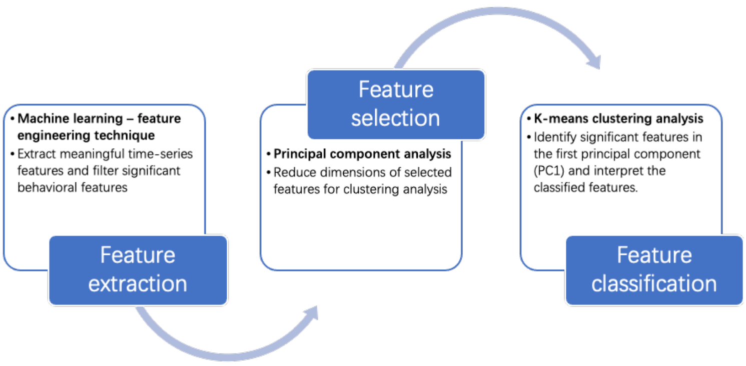



**Figure 1. Analysis procedure**

# 3 RESULTS

As a result, we extract 787 time series features and sort out 178 features that are highly related to learning using the machine learning technique. We further implement PCA for dimension reduction to 20 out of 178 significant time-series features indicating employees' learning behaviours for clustering analysis. Leveraging PCA and K-Means clustering, we find the model with k = 3 clusters exhibits robust clustering performance, as evidenced by a relatively high Calinski-Harabasz Score, signifying well-separated clusters, and a Silhouette Score suggesting some degree of cluster separation.

Specifically, we identify three significant behavioral features among clusters. These features include approximate entropy (ApEn), permutation entropy (PeEn), and sample entropy (SampEn, Ferrario et al., 2005), indicating employees' regularity of learning time. ApEn assesses the level of irregularity or randomness within a given data series. It accomplishes this by quantifying the probability that similar patterns will not recur in the data. This is achieved through the comparison of sequences of data points, allowing us to identify both similarities and differences. Similarly, SampEn serves as a tool to evaluate the regularity and predictability of data series. It quantifies the likelihood of observing similar data sequences within the data. PeEn focuses on the unpredictability of patterns within a time series dataset. It does so by comparing permutations of data points within defined windows. A higher entropy value in any of these measures indicates a greater degree of irregularity in the data, implying a lower predictability and a more intricate structure.

**Table 1. K-mean clustering indices of selected variables**

| Group | N | Mean score | RM_rank_perc | Count_dorm | ApEn | PeEn | SampEn |
|---|---|---|---|---|---|---|---|
| 0 | 469 | 43.59 | 0.47 | 14.38 | 0.79 | 5.37 | 1.78 |
| 1 | 738 | 48.01 | 0.46 | 17.42 | 0.89 | 5.41 | 2.08 |
| 2 | 427 | 54.07 | 0.45 | 22.67 | 0.64 | 5.21 | 1.12 |

The ANOVA test compares the variance across the means of the identified groups (Table 1). There are significant differences in behavioral features, {$F_{ApEn}(2, 1632) = 612.05$, $F_{PeEn}(2, 1632) = 216.60$, $F_{SampEn}(2, 1632) = 522.79$, $p_s < .001$}, learning outcomes, F (2, 1632) = 3.45, $p < .001$ , learning efficiency, $F$ (2, 1632) = 76.61, $p < .001$, and business performance, $F$ (2, 1632) = 4.23, $p = .01$ among the three groups of individuals. Specifically, post-hoc tests show individuals in Cluster 0 exhibit the lowest learning and business performance. They engage in more practice sessions, indicating their commitment to skill improvement. However, they display less consistency in their learning, suggesting significant variations in session timing. In contrast, individuals in Cluster 1 demonstrate moderate learning and business performance. Meanwhile, individuals in Cluster 2 participate in the least number of practice but maintain the highest correctness rate. Additionally, they exhibit a greater degree of regularity in their learning time patterns. Consequently, individuals in Cluster 2 achieve the highest learning and business performance with highest self-paced training efficiency.

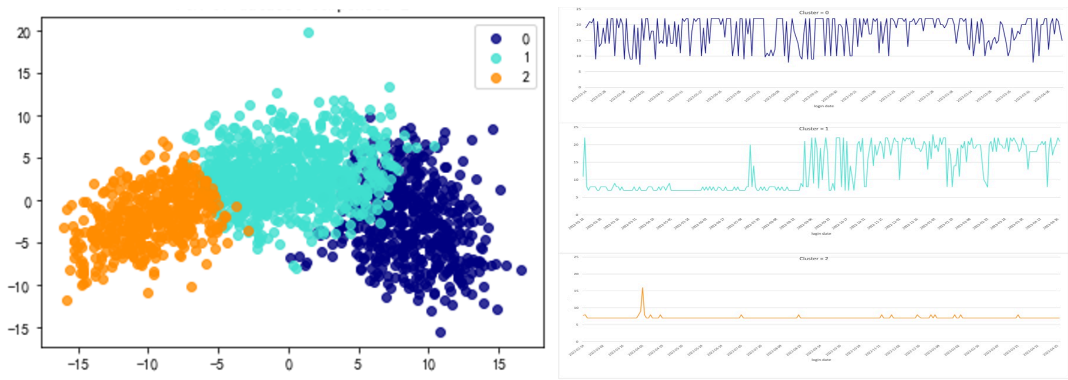



**Figure 1. A comparison between regularity patterns across clusters**

Figure 1 illustrates the clustering results (left) and the finer grained of the learning patterns (right). The y-axis (right) represents employees' active time from 7am to 23pm during the period of training program, and the x-axis represents the duration of starting date to the completion date across fourteen months. It shows that individuals in Cluster 0 demonstrate the widest range of learning times (i.e., highest irregularity) during the training period. While individuals in cluster 1 reveals similar patterns of learning; initially, they are more regular at the beginning of the professional development program, but as the program continued, they do not maintain the same level of learning regularity and demonstrate more random learning patterns. In contrast, individuals in Cluster 2 are able to maintain relatively consistent learning times (highest regularity).

## 4 DISCUSSION

Our study aims to identify effective learning patterns in to improve training efficiency. Our research unveils: employees who maintain regular learning schedules benefit the most, both in professional skill development and business performance. The findings have practical implications for optimizing training strategies and emphasizing time management in training programs. Furthermore, our research provides empirical evidence for using learning analytics in workplace online training, enhancing our understanding of employee learning patterns in organizational contexts. In sum, our study addresses the value of learning analytics in understanding employees' learning behavior and offers practical insights for improving workplace training efficiency and organizational growth.

## REFERENCES

Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics,* 2(4), 433-459.

Christ, M., Braun, N., Neuffer, J., & Kempa-Liehr, A. W. (2018). Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package). *Neurocomputing*, *307*, 72-77.

Ferrario, M., Signorini, M. G., Magenes, G., & Cerutti, S. (2005). *Comparison of entropy-based regularity estimators: application to the fetal heart rate signal for the identification of fetal distress. IEEE Transactions on Biomedical Engineering,* 53(1), 119-125.Australian Psychological Society. (2008).

Wang, X., Lin, W., Xue, T., Green, A., Gu, L., He, Y., ... & Wu, Y. (2022). Workplace learning in China: Transferring training into practice to improve performance. *International journal for research in vocational education and training*, *9*(3), 294-315.

# The Path to AI SME: Harnessing Learning Analytics to Enlighten Conversational Agents for Self-Directed Study

**Author: Mu He**
Canadian Security Institute
Jason.He@moodys.com

**ABSTRACT**: This paper presents an approach to improving large language model chatbots using learning analytics. The authors developed an AI subject matter expert to support self-directed learners. To reduce hallucinated responses, they utilized embeddings from course materials to constrain the model outputs. By leveraging granular learning progress data, the authors optimized embedding retrieval to only extract the most relevant documents. Tests showed the context-aware procedure attained higher response rates and lower costs than the conventional RAG approach, especially with a large vectorstore. The results suggest integrating analytics into these systems can enhance adaptivity and reduce cost. Further development is still needed, but this demonstrates promise for analytics-informed conversational agents in adult learning.

**Keywords**: AI, Adult learning, Student Support, Learning Analytics, Self-directed Learning,

## 1 BACKGROUND

The Canadian Securities Institute (CSI) offers education programs and credentials in Canada and around the world. Our goal for this project is to transform self-directed online adult learning through personalized and adaptive learning driven by artificial intelligence. By integrating cutting-edge large language models, we developed a virtual subject matter expert (SME) to provide personalized learning support that helps our online learners from anywhere and at any time.

In our pursuit of this goal, we have encountered a range of challenges in this project. Chief among them is the issue of hallucinations in Gen-AI models and the imperative need to ensure cost-effectiveness. To address these hurdles, we have harnessed the potential of learning analytics, leveraging its insights to devise effective solutions.

## 2 IMPLEMENTATION

Large language models (LLM) like the GPT series from OpenAI can sometimes generate hallucinated or fabricated responses that seem convincing but are not grounded in facts (Kenton et al., 2021). One way to reduce this tendency towards hallucination is to use embeddings. Embeddings are vector representations of words, phrases, or larger bodies of text that encode semantic meaning. By combining embeddings and prompt engineering with the OpenAI models, the generations become more constrained to reflect actual knowledge contained in the embeddings (Coenen et al., 2019).

In the development of our AI SME, we utilize embeddings derived from key materials like the textbook, practice quizzes, and other references to create a series of domain-specific vectorstores. However, since the volume of the embeddings is large, querying across all vectorstores to locate pertinent embeddings for a given question can be time-consuming and expensive (Chen et al., 2020). Additionally, if the number of returned documents (K value) is not large enough, the LLM may fail to identify useful target embeddings from the corpus which leads to a low response rate (Thoppilan et al., 2022).

To address these challenges, we leveraged our multi-dimensional learning analytics system which contains rich student learning data gathered from the Learning Records Store (LRS), Learning Management System (LMS), and video analytics platform. With the chapter-level or even page-level learning progress data, we can pinpoint the most relevant embedding sets to include for a given question, rather than querying across all available vectorstores. This allows us to provide the LLM with additional context to improve the semantic search results and to reduce the chance of hallucination. Moreover, the optimized embedding retrieval informs language model generations that directly align with the most relevant knowledge and avoid including the contents or concepts that student has not encountered. For example, the learner profile data can reveal our learner is studying the content in Chapter 2. We can then pass it as a parameter to query vectorstores in an optimized way to extract embeddings before Chapter 3 as well as other supporting materials.

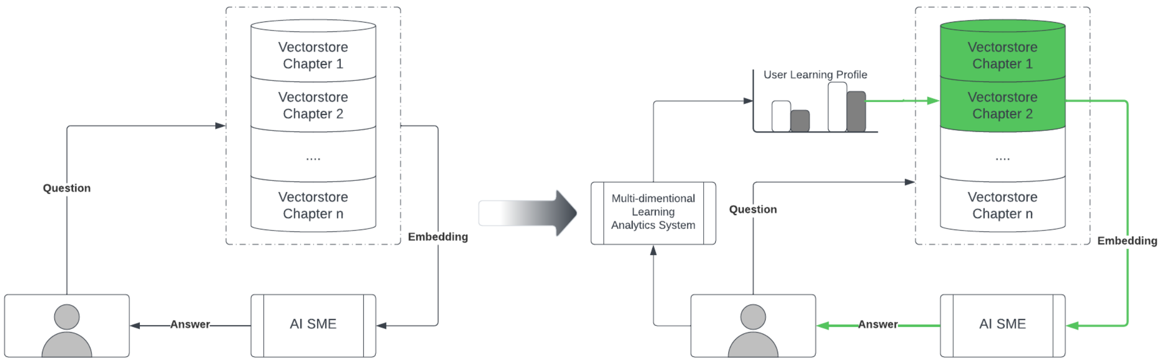



**Figure 1: Using learning analytics to improve the AI SME outputs.**

## 3 FINDINGS

We systematically compiled a corpus of 66 genuine student questions sourced from 27 course discussion forums that had been previously addressed by human SMEs. These questions were annotated with labels indicating the topic, subject matter, associated textbook chapter, and other relevant attributes. We then fed these questions into two procedures - one that received the associated chapter as a parameter, and one that did not. Both procedures utilized the same underlying large language model, temperature, retriever, and system message. By comparing the performance of the context-aware versus conventional procedures, we aimed to quantify the differences between the two approaches giving different numbers of documents to return (K value).

**Table 1: Comparison of Context-aware and Conventional RAG Procedures when K equals to 5.**

| Procedure | Avg. Time | Avg. Token | Avg. Cost | Response Rate |
|---|---|---|---|---|
| Context-aware | 12.02s | 1793.78 | 0.1281¢ | 87.88% |
| Conventional | 11.57s | 1814.38 | 0.1285¢ | 75.76% |

**Table 2: Comparison of Context-aware and Conventional RAG Procedures when K equals to 10.**

| Procedure | Avg. Time | Avg. Token | Avg. Cost | Response Rate |
|---|---|---|---|---|
| Context-aware | 13.44s | 2757.57 | 0.1868¢ | 96.97% |
| Conventional | 13.16s | 2788.12 | 0.1877¢ | 86.36% |

**Table 3: Comparison of Context-aware and Conventional RAG Procedures when K equals to 25.**

| Procedure | Avg. Time | Avg. Token | Avg. Cost | Response Rate |
|---|---|---|---|---|
| Context-aware | 14.97s | 5580.29 | 0.3567¢ | 100% |
| Conventional | 14.33s | 5759.38 | 0.3671¢ | 96.97% |

The above results demonstrate that, for a given K value, the context-aware procedure consistently outperforms with higher response rates and lower costs. Most notably, the context-aware approach achieves the same response rates at K=10 as the conventional approach at K=25. In this scenario, the context-aware procedure is faster while significantly reducing costs by about half. Moreover, the context-aware procedure generates responses based on content the learner has already studied, while the conventional approach may refer to content the learner has not yet encountered. In some cases, this could lead to more confusion than clarification. These findings underscore the advantages of integrating learning analytics data to improve the performance of LLM-based AI chatbot applications.

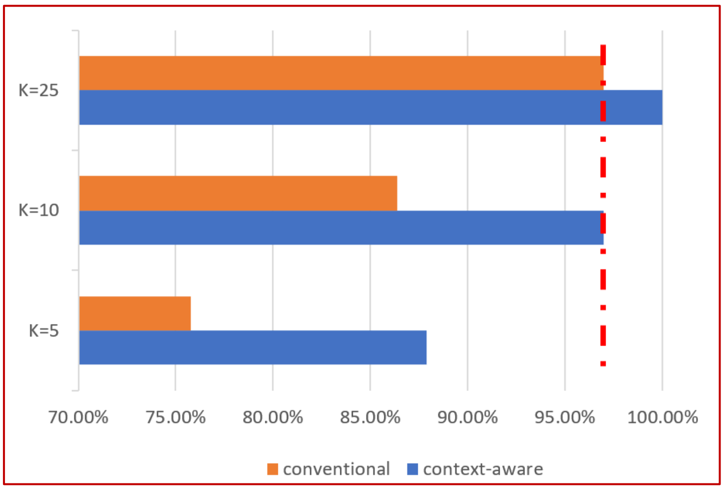



**Figure 2: The context-aware approach can achieve high response rate with smaller K value.**

# 4 DISCUSSION

We presented an innovative approach to integrating insights from a multi-dimensional learning analytics system into an LLM-based virtual AI SME. To evaluate the performance, we conducted multiple rounds of evaluation with different numbers of document retrieval that showed promising results. As suggested by prior work, incorporating learning analytics into intelligent systems can enhance the personalization and adaptivity of the system (Ifenthaler & Widanapathirana, 2014). Future work should focus on incorporating additional metrics from the learner profile, such as motivation levels, and progress on practice exercises, to further refine the embedding retrieval process. Furthermore, feeding the learner's chat history back into the learning analytics system could enable continuous improvement of the system's adaptivity and personalization.

While our approach demonstrates promising initial steps toward integrating learning analytics in LLM-based applications, further rigorous learner-centered development and testing are required to fully realize the potential. The results presented verify the advantages of leveraging learning analytics to significantly enhance the performance and reliability of conversational agents. However, continued research into optimizing the learning analytics integration, utilizing multimodal LLMs, and gathering human feedback will be critical to the goal of creating truly adaptive, useful, and human-like AI SMEs.

## REFERENCES

Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P., Kaplan, J., Edwards, H., Burda, Y., Joseph, N., Brockman, G., Ray, A., Puri, R., Krueger, G., Petrov, M., Khlaaf, H., Sastry, G., Mishkin, P., Chan, B., Gray, S., . . . Zaremba, W. (2021). Evaluating Large Language Models Trained on Code. ArXiv. /abs/2107.03374

Coenen, A., Reif, E., Yuan, Z., Kim, B., Pearce, A., Viégas, F., & Wattenberg, M. (2019). Visualizing and measuring the geometry of BERT. In Advances in Neural Information Processing Systems.

Ifenthaler, D., Widanapathirana, C. Development and Validation of a Learning Analytics Framework: Two Case Studies Using Support Vector Machines. Tech Know Learn 19, 221–240 (2014). https://doi.org/10.1007/s10758-014-9226-4

Kenton, Z., Everitt, T., Weidinger, L., Gabriel, I., Mikulik, V., & Irving, G. (2021). Alignment of Language Agents. ArXiv. /abs/2103.14659

Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H., Jin, A., Bos, T., Baker, L., Du, Y., Li, Y., Lee, H., Zheng, H. S., Ghafouri, A., Menegali, M., Huang, Y., Krikun, M., Lepikhin, D., Qin, J., . . . Le, Q. (2022). LaMDA: Language Models for Dialog Applications. ArXiv. /abs/2201.08239

# "Close...but not as good as an educator" - Using ChatGPT to provide formative feedback in large-class collaborative learning

**Cory Dal Ponte, Sathana Dushyanthen, Kayley Lyons**
Centre for Digital Transformation of Health, University of Melbourne
cory.dalponte@unimelb.edu.au

**ABSTRACT**: Delivering a personalised learning experience that includes real-time data collection, data collection and feedback has been difficult without using a custom-built analytical tool. We employed ChatGPT as a scalable, novel, real-time analytical tool in a one-hour Zoom break-out room activity that taught practicing health professionals how to formulate evaluation plans for digital health initiatives. Learners completed an evaluation survey that included Likert scales and open-ended questions that were analysed. Half of the 44 survey respondents had never used ChatGPT before. Overall, respondents found the use of the tool favourable, described a wide range of group dynamics, and had adaptive responses to the feedback. Future educators can learn from our experience including engineering prompts, providing instructions on how to use ChatGPT, and scaffolding optimal group interactions with ChatGPT. Future researchers should explore the influence of ChatGPT on group dynamics and derive design principles for the use of ChatGPT in collaborative learning.

**Keywords**: Collaborative learning, ChatGPT, generative AI, real-time analysis, formative feedback

## 1 INTRODUCTION

In blended learning, collaborative problem-based learning frequently serves as a method for having students apply their understanding acquired in self-directed learning modules (Hmelo-Silver & DeSimone, 2013). Ideally, each problem-based learning group would have a facilitator supporting and providing valuable feedback, checking for misunderstandings, and encouraging discussion. However, similar to many courses, in our Applied Learning Health Systems professional development course, we only have two instructors for the ten groups completing activities in virtual break-out rooms. Class-wide feedback is provided in debriefs after the activity; however, this feedback does not provide specific feedback for each group.

We introduced a novel learning analytics approach using ChatGPT to personalise feedback within a large class breakout room activity. Using ChatGPT shifts the analytical process to the moment of learning rather than afterwards, providing real-time data collection, analysis, and feedback on group output (Shimada, Konomi and Otaga, 2018), making the information immediately useful to students (see Figure 1). The tool could have the potential to influence group dynamics and perceptions of learning as well as improving output prior to a large class debrief.

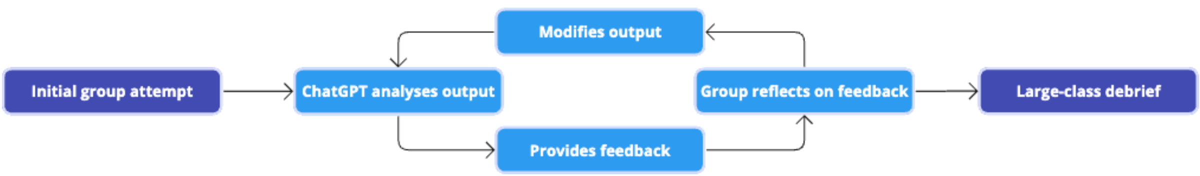


**Figure 1. Real-time analysis and feedback loop.**

To evaluate the potential benefits and unintended consequences of integrating this tool, we posed the following research questions:

1. How did participants perceive the quality of the feedback provided by ChatGPT?
2. How did participants describe the impact on their learning and group dynamics?
3. What reasons do students attribute to the quality ratings (e.g., Beginner, Advanced) received from ChatGPT on their evaluation plans?

## 2 METHODS

We used the computer-supported collaborative learning (CSCL) design framework proposed by Zheng (2021) to describe our activity and integrate ChatGPT for real-time analysis and feedback. The course was a 13-week professional development course for practicing health professionals. Each week a different digital health topic (e.g., implementing digital solutions) is presented based on the Learning Health Systems framework. The activity was piloted during one week of the course. The goal for this week was to enable participants with no prior research training to craft a concise evaluation plan that incorporated fundamental components and ensured alignment with the research question. Before the workshop, participants completed three hours of self-directed modules. Then, in the 2.5-hour workshop, the ChatGPT activity was completed in a 45-minute Zoom® breakout room activity with groups of 5-7 participants. Prior to the activity, we provided a 15-minute ChatGPT demonstration and detailed step-by-step instructions. Groups had 25 minutes to craft their evaluation plan followed by 20 minutes for participants to gather and react to ChatGPT feedback.

We've included a link to view the conversation with GPT4 to engineer the instruction and prompts. As ChatGPT requires structured instructions to get the desired output, we engineered the following set of standardized custom instructions and prompts for students to use within the activity (see Figure 2):

Act as an expert in the learning health systems framework with a focus on socio-technical evaluation plans. When provided with an evaluation plan table, meticulously analyse its content based on the specific categories:

1. **Method of Evaluation**: Assess the appropriateness, robustness, and feasibility of the chosen evaluation method.
2. **Theoretical Framework**: Ensure that the theoretical framework is relevant to the evaluation's objectives and is applied correctly. If the plan does not mention a theoretical framework then propose an appropriate theory based on the objectives.
3. **Justification for Evaluation and/or Theoretical Framework**: Critique the provided justification, ensuring it's coherent and aligns with the chosen evaluation method and theoretical framework.
4. **Data Collection Methods**: Evaluate the chosen data collection methods for their relevancy, comprehensiveness, and potential biases.
5. **Data Sources**: Critically examine the listed data sources for their relevance, reliability, and potential to address the evaluation's objectives.
6. **Brief Description of Data Analysis**: Analyse the clarity, comprehensiveness, and relevance of the data analysis description. Ensure it fits the context of the evaluation and theoretical framework. If it's missing, then propose a data analysis plan.

Provide both strengths and potential areas of improvement for each plan, ensuring feedback is constructive, clear, and actionable.

I'd like to receive feedback on the socio-technical evaluation plan below.
Please provide constructive feedback on the criteria and offer suggestions for potential areas of improvement. Provide a rating from beginner to advanced on the overall evaluation plan based on how well it addresses the rationale and objectives.

Rationale and objectives: *your team's answers here*
Method of Evaluation: *your team's answers here*
Theoretical Framework: *your team's answers here*
Justification for Evaluation: *your team's answers here*
Data Collection Methods: *your team's answers here*
Data Sources: *your team's answers here*

**Figure 2. ChatGPT Custom Instructions and Prompts**

This study was approved by the University of Melbourne human ethics committee (Project ID 22641). Following the activity, participants immediately completed a Qualtrics® survey. The survey included a mix of open-ended questions and Likert scales according to the research questions. Likert scales were presented as descriptive statistics. Open-ended questions were analysed by a qualitative researcher who open-coded responses according to the research questions. After themes were determined by open-coded, we conducted a frequency analysis of themes.

# 3 RESULTS

The session included 55 participants from 13 organisations (92% hospitals, 8% University) and 18 job titles (45% health professions, 18% health services management, 13% researchers, 18% IT/data analytics, 5% consumers). 44 participants (80%) completed the survey. Half of the participants (51%, n=23) had never used ChatGPT before, with 27% (n=12) using it once a month, 16% (n=7) weekly and 5% (n=2) daily. Through an analysis of the groups' conversation record with ChatGPT, we identified that the majority (7 out of 10) of groups did not iterate their evaluation plans after receiving one round of feedback. We asked participants how they perceived the quality of feedback they received (see Table 1.)

**Table 1: Perceived quality of feedback received from ChatGPT (n=44)**

| Field | Strongly agree | Agree | Neither agree nor disagree | Disagree | Strongly disagree |
|---|---|---|---|---|---|
| The ChatGPT feedback was valuable | 32% | 52% | 11% | 5% | 0% |
| The ChatGPT feedback helped me learn about digital health evaluation | 16% | 52% | 20% | 11% | 0% |
| The ChatGPT feedback influenced us to improve our evaluation plan | 30% | 55% | 14% | 2% | 0% |
| The ChatGPT feedback increased my interest in digital health evaluation | 16% | 32% | 36% | 16% | 0% |
| The ChatGPT feedback increased my interest in generative AI | 45 % | 41% | 11% | 2% | 0% |

*Influence on learning*: In response to the question "How did the personalised feedback from ChatGPT influence your learning about digital health evaluations? If anything?," most participants wrote that ChatGPT provided specific (n=12), actionable (n=9), immediate (n=7), and easy to understand (n=3) feedback on their evaluation plans. Six participants wanted the feedback validated by a human before deciding to learn from the feedback, or noted they did not have the expertise to judge the quality of the output. Exemplifying the mixed comment group was one participant who noted that the feedback was "close but not as good as an educator." Negative written comments (n=6) either simply stated ChatGPT did not influence their learning, or the feedback was too vague.

*Influence on group dynamics*: We found a wide variety of written responses regarding how ChatGPT influenced group dynamics. Twenty-four participants noted how it enhanced group discussion due to the group responding to the feedback (n=16), becoming behaviourally engaged due to the novelty of ChatGPT (n=4), and ChatGPT activating the discussion (n=3). Six participants wrote that ChatGPT did not influence group dynamics. Surprisingly, 14 participants described how ChatGPT hindered their group dynamics either by providing a distraction, causing group members to spend much more time reading the output, or providing an authority within the group. For example, one participant noted, "ChatGPT essentially acted as the final arbiter." Another participant described that it hindered discussion because "with ChatGPT in the room, there was less need for general discussion as we could simply ask it for answers."

*Causal attributions*: In response to the survey question, "Why do you think your group received either a beginner, intermediate, or advanced rating from ChatGPT?," rarely did participants not trust or disagree with their rating. In terms of causal attribution theory (Kelley, 1973), 31 participants gave descriptions that were coded as adaptive responses to feedback (i.e., internal, non-stable, in-control cause, e.g., "we didn't give enough information on what the data analysis part looked like in-depth"). Counter to our expectations, only one participant distrusted their GPT rating. This participant said their rating was received "because it [ChatGPT] does not know enough."

# 4 DISCUSSION & RECOMMENDATIONS

Overall, we demonstrated favourable perceptions from a group of learners who were mostly new to ChatGPT and practicing professionals. ChatGPT facilitated a real-time analytical feedback loop within the learning activity, a reality that has been limited in collaborative learning environments without building custom tools that limit their scalability. Part of the increased engagement and interest in the activity could be explained by the novelty effect of new technology although a few students mentioned this as a distraction.

In future iterations of this activity, we aim to improve the specificity of the ChatGPT feedback through further prompt engineering. As many groups' answers were lacking in detail due to the time constraints, we will increase the focus of prompts on suggesting examples on how to improve their inputs. Also, we will enhance the structure of the activity to ensure the learners return to their evaluation plans and improve them.

We used OpenAI ChatGPT 3.5 interface with equity and scalability in mind. Future educators can learn from the design of our activity and prompts. Based on our findings, for educators integrating generative AI (genAI) into large class size collaborative learning environments, we recommend:

1. Engineer and rigorously test standard custom instructions and prompts
2. Provide detailed instructions outlining how students should interact with the tool
3. Scaffold how the group should optimally interact with ChatGPT

Future research should focus on identifying sets of design principles that assist educators to optimally utilise genAI as an analytical tool. Given the diversity in group dynamics in our study, future researchers should observe and investigate the factors for why some groups had rich discussions while others had superficial or no discussion at all. Also, researchers should continue to explore student beliefs with using genAI.

# REFERENCES

Hmelo-Silver, C.E. & DeSimone, C. (2013). Problem-based learning: An instructional model of collaborative learning. In *The international handbook of collaborative learning*. Routledge, pp. 370-385.

Shimada, A., Konomi, S. and Ogata, H. (2018), Real-time learning analytics system for improvement of on-site lectures, *Interactive Technology and Smart Education*, Vol. 15 No. 4, pp. 314-331. https://doi.org/10.1108/ITSE-05-2018-0026

Kelley, H. H. (1973). The processes of causal attribution. *American psychologist*, *28*(2), 107.

Zheng, L. (2021). An Innovative Framework for Designing Computer-Supported Collaborative Learning. In: Data-Driven Design for Computer-Supported Collaborative Learning. Lecture Notes in Educational Technology. Springer, Singapore. https://doi.org/10.1007/978-981-16-1718-8_1

# Evaluating the Use of Canvas' LMS New Analytics Tool in Language Course Engagement and Design: A Case Study from a Hong Kong University Language Center

**Author(s): Mansurbek Kushnazarov**
The Hong Kong University of Science and Technology
lcmansur@ust.hk

**Author(s): Delian Gaskell**
The Hong Kong University of Science and Technology
lcdgaskell@ust.hk

**ABSTRACT**: This **practitioner paper** evaluates efforts to introduce and support the use of New Analytics, a built-in function of Canvas LMS, to highlight levels of student engagement with various learning materials for language teachers and subsequent need for amendments to course design in a university language center in Hong Kong. The teachers received bi-weekly reports of their course's student engagement and participation data in the form of written summary descriptions, tables, and graphs. The aim was to promote the use of New Analytics as a practical tool to introduce learning analytics (LA) from a 'bottom up' approach to teachers and to provide opportunities for evidence-based course revision since this tool was readily available to all Canvas users. However, the authors faced several challenges in implementing the outcomes for the initiative, such as increased workload, data formatting, material grouping, and data interpretation and application, and resistance to the concept of 'tracking' from a vocal minority of colleagues. The paper discusses the limitations and implications of adopting Canvas' New Analytics tool and offers recommendations on implementing the same or similar commercial LA tools.

**Keywords**: New Analytics, language education, course design, learning materials engagement, blended learning

## 1 INTRODUCTION

The integration of learner engagement with learning goals, tasks, and context is conducive to effective language learning (Zhang, 2020). Though it could be debated whether this enhances or hinders students' engagement, learning management systems (LMS) often serve as repositories of educational material or platforms for interactive learning objects (Holmes & Prieto-Rodriguez, 2018). Meanwhile, learning analytics (LA) provides access to authentic learning behaviour data that is free from subjectivity, enabling learning designers and teachers to make informed decisions regarding teaching and learning (Khalil et al., 2022). In Higher Education, LA also tends to focus on supporting students by guiding them to become self-regulated and self-directed learners (Viberg et al., 2020). The university where this project took place has not been an earlier adopter of Blended Learning as a mainstream pedagogical approach. While Blended Learning approaches have the potential to enhance engagement and interactivity in LMS materials, this study found that teachers were initially reluctant to adopt this approach unless it was mandated. Furthermore, when courses were identified for the transformation into Blended Learning materials, the writers and designers encountered laborious and challenging tasks. Once the materials were 'finished,' they lacked motivation for further iteration and were uninspired to make interactive improvements. In this prevailing culture, course writers and teachers lacked the knowledge and skills to collect and analyze data on student engagement and learning outcomes, hindering their ability to understand whether the materials were being used by the target students, and whether these materials were being used interactively. To address these issues, this study aimed to introduce colleagues at a Hong Kong university language center to Canvas'

New Analytics features. We were informed by the literature that LA requires a flexible approach (Ifenthaler, 2020) to spot successful learning patterns (Crossley et al., 2016), to identify learning misconceptions and misplaced effort (Poitras et al., 2016), and to implement appropriate interventions within our colleagues' courses since these New Analytics tools were intended to provide insight into and measurement of students 'engagement' with the learning material. The study's results, from qualitative interviews with participating teachers and reflections from the project leaders, identified opportunities and challenges associated with implementing a commercially available LA tool.

## 2 CONTEXT

The strategy of adopting Canvas's New Analytics[1] tool to inform course and learning materials design took place in a university language center in Hong Kong. The project leaders invited language teachers to participate in our pilot study by nominating a Blended learning course with interactive learning content (e.g., video, SCORM, 3rd party apps). Table 1 illustrates the details of the courses that participated in Phases 1 and 2.

**Table 1: Details of courses participating in Phases 1 and 2.**

| Phase | Course | Focus/level | Number of students |
|---|---|---|---|
| Phase 1 (Fall 2022) | English | Science communication (Math) | 20 |
| | English | Common Core, academic English | 20 |
| | Chinese | Common Core, elementary Putonghua | 20 |
| Phase 2 (Spring 2023) | English | Common Core, academic English | 17 |
| | English | Common Core, Digital Literacy | 57 |
| | Chinese | Common Core, elementary Putonghua | 18 |
| | Chinese | Common Core, elementary reading and writing | 17 |
| | Japanese | Beginner | 20 |

The project leaders extracted the weekly online activity data from New Analytics for each course and made bi-weekly reports with three sections: 1) student engagement with learning materials, 2) views of each learning material, and 3) participation (i.e., forum posts). Each section had a table, a line graph, and a summary of the main trends to help course teachers understand the interaction over two weeks.

The learning materials of the courses were categorized into four groups based on Law and Liang's (2020) framework: 1) interactive exercises, 2) static texts, 3) discussion forums, and 4) assignments. Each group had a different color in the reports to help identify and describe engagement patterns. The course teachers got bi-weekly reports on student engagement on Miro boards. The reports aimed to help teachers improve their teaching, materials, and assessments during and after the semester. In Phase 1, the reports were delayed by 7-10 days due to the time needed for preparation. In Phase 2, the reports were given only at the end of the semester, reducing the teachers' chance to use the data. This change was justified by two reasons: first, Phase 1 teachers said they had no time

[1] New Analytics in Canvas is a tool that allows users to track students' interactions and participation in visual and tabular formats and make informed decisions about course design and student learning. It enables instructors to track average course grades for student submissions and view page views and participation metrics across all devices.

to make changes based on the reports. Second, the project leaders had more duties in the department, hindering them from making timely reports.

## 3 RESULTS: BENEFITS AND CHALLENGES OF IMPLEMENTING NEW ANALYTICS TO INFORM COURSE DESIGN

The course teachers were interviewed after Phases 1 and 2. The qualitative data was analyzed using content analysis. The teachers appreciated the reports that showed them how students engaged with the learning materials. They said the reports gave them insights and questions for course improvement. For instance, a teacher noticed an article that had 360 views from 57 students, and decided to check its difficulty level. Another teacher used the reports to improve self-directed learning materials for future student engagement. The teachers also wanted clearer instructions on how to read the reports and use the insights for course revisions. Although New Analytics was supposed to be user-friendly and intuitive, and some teachers found it easy to use, the tool and our support did not significantly affect the use of more interactive materials or course design. The results showed passive resistance to technology and course design in the department. This resistance came from various factors, such as pedagogical beliefs that did not value technology, worries about tracking, and policies that gave teachers freedom in teaching. The teachers were also reluctant to revise large Blended Learning courses because they had already spent a lot of time and manpower creating them (Kaliisa et al., 2020). Simultaneously, the project leaders faced challenges with data conversion, interpretation, and categorization. These challenges limited the potential of LA data to inform course design for better student engagement.

## 4 INSIGHTS AND IMPLICATIONS

This study demonstrated how a bottom-up approach with a commercial LA tool, New Analytics, could enhance course and learning materials design at a language center. The authors also gained insights that could benefit other practitioners in similar contexts. Based on the findings, they offer suggestions for using LA tools like New Analytics in a department of a large research-oriented university and beyond.

1. Begin with a small-scale experiment in a single course to gain a deeper understanding of the advantages and constraints associated with utilizing the LA tool to inform course design. Evaluate the workload necessary to gather detailed data and transform it into concise written descriptions and charts. This will help determine the feasibility and effectiveness required for timely and reflective adjustments to the learning materials.
2. Demonstrate the project's next steps to interested teachers, showing how to use the tool's data to design and improve courses. Compare the feedback from this data and self-reported methods to evaluate their quality and validity. This will help teachers to understand how effective a LA tool can be to obtain course feedback and make course design decisions.
3. Train teachers to use New Analytics effectively, covering navigation, functions, charts, and limitations.
4. Use a framework (e.g., Law & Liang (2020)) to analyze and classify different kinds of learning materials in a course. Grouping learning materials helps to understand how learners engage with different types of materials and allows for a deeper analysis of learner engagement. The teachers in our project found this practice enlightening.

5. Explore how new GenAI can help analyze and interpret data visualizations. Share the learned skills with teachers, explaining how GenAI can enhance their understanding of student engagement with online learning materials, and make informed decisions to improve the learning experience in and out of classrooms.
6. Invite teachers who have used New Analytics successfully to share their insights and experiences with others. Organize workshops, seminars, or blogs to facilitate knowledge sharing and discussions. This can help more teachers see the benefits and challenges of using LA tools and change their perception of such technologies from doubt to acceptance.

## 5 CONCLUSION

The adoption of Canvas' New Analytics or similar built-in LMS LA tools have potential benefits but also present several challenges. These tools must be designed with an understanding of teachers' workflows and needs in mind to maximize their utility. Further support from relevant university departments for credibility and wider scope of reach, research and peer sharing are needed to guide the development and implementation of more effective LA tools in language teaching settings, and beyond. While this study is based on the course teachers' interviews and the project leaders' reflections, more data from other stakeholders such as students and curriculum developers are needed to enhance the validity of the findings. Other limitations of this study include the lack of comparison of the New Analytics tool from Canvas with other commercial LA platforms and LMSs and the project length that encompasses a single academic year. The authors aim to remediate such constraints in subsequent phases of the study, given approval and support by the university department.

## REFERENCES


Crossley, S., Paquette, L., Dascalu, M., McNamara, D. S., & Baker, R. S. (2016). Combining click-stream data with NLP tools to better understand MOOC completion. Proceedings of the *Sixth International Conference on Learning Analytics & Knowledge*, Edinburgh, United Kingdom.

Holmes, K., & Prieto-Rodriguez, E. (2018). Student and staff perceptions of a learning management system for blended learning in teacher education. *Australian Journal of Teacher Education (Online)*, 43(3), 21–34.

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.

Kaliisa, R., Kluge, A., & Mørch, A. I. (2020). Combining Checkpoint and Process Learning Analytics to Support Learning Design Decisions in Blended Learning Environments. *Journal of Learning Analytics*, 7(3), 33-47.

Khalil, M., Prinsloo, P., & Slade, S. (2022). A Comparison of Learning Analytics Frameworks: a Systematic Review. *LAK22: 12th International Learning Analytics and Knowledge Conference*.

Law, N., & Liang, L. (2020). A multilevel framework and method for learning analytics integrated learning design. *Journal of Learning Analytics*, *7*(3), 98-117.

Poitras, E., Naismith, L., Doleck, T., & Lajoie, S. (2016). Using Learning Analytics to Identify Medical Student Misconceptions in an Online Virtual Patient Environment. *Online Learning*, 20.

Viberg, O., Kalili, K., & Baars, M. (2020). Self-regulated Learning and Learning Analytics in Online Learning Environments: a Review of Empirical Research. *LAK20: 10th International Learning Analytics and Knowledge Conference*.

Zhang, Z., (2020). Learner engagement and language learning: a narrative inquiry of a successful language learner. *The Language Learning Journal, 50*:2, 378-392.

# The Data Hub of the Institute for the Future of Education

**Héctor G. Ceballos, Gerardo Castañeda, Joanna Alvarado-Uribe, Paola Mejía-Almada and Ubaldo Martinez-Carranza**
Tecnologico de Monterrey, Institute for the Future of Education
{ceballos, g.castaneda, joanna.alvarado, gabriela.almada, ubaldo.martinez}@tec.mx

**ABSTRACT**: This manuscript presents the experiences and lessons learned from the first two years of the Data Hub of the Institute for the Future of Education. This initiative tackles several issues commonly faced by Higher Education Institutions (HEIs): integration of educational data, securing data privacy, data curation, business rules documentation, providing a single point to obtain up-to-date validated educational information, and building a research community (through open data-driven calls, workshops, and webinars). This initiative provides access to granular educational data from multiple sources involving all the necessary stakeholders in the process. We provide a basis for the design of an educational Data Hub that serves as an enabler for Learning Analytics research.

**Keywords**: Educational data hub, data acquisition, data curation, data governance, data science, educational innovation, learning analytics, higher education institutions.

## 1. INTRODUCTION

Following the trend of the new possibilities that data academic generation and integration provides over the last two decades, Higher Education Institutions (HEIs) have the opportunity to be more aware of the relevance of their data. For Learning Analytics, the availability of this data represents an opportunity to review institutional strategies in favor of better understanding their learning outcomes – and moreover, their areas for improvement. Nonetheless, challenges arise as different types of data are stored across different departments, with information related to academic records that may not be managed by the same people and department (e.g. information related to surveys, graduate records, continuing education, etc.).

Additionally, over time, storage, documentation, and updating of this data may not be adequate for research and decision-making purposes: changes in systems/software, file formats, or change in subject matter experts leave organizations with difficulties in comprehending the real value and meaning of their data. In consequence, while the source of the data remains valid, it still requires further processing and review before being suitable for internal or external research use.

## 2. THE DATA HUB OF THE INSTITUTE FOR THE FUTURE OF EDUCATION

In 2019, our Institution founded the Institute for the Future of Education[1] through the consolidation of multiple initiatives on educational innovation. Our institute has for mission to improve the life of millions of people around the world through higher education and lifelong learning. Its initiatives include: funding educational innovation projects, conducting experimental, basic and applied

[1] Institute for the Future of Education (IFE). Tecnológico de Monterrey. https://tec.mx/en/ife

research (Research Lab), promoting open innovation, sharing best practices through an educational innovation observatory, and, among others, constituting an educational data repository for research.

Having a long trajectory in educational research, the Institute for the Future of Education's leaders identified the inaccessibility to granular educational data as one of the main obstacles to conducting quantitative education research. For this reason, in December 2021, the Data Hub was officially launched and presented as a space to provide real, curated, and anonymized educational data to researchers from the institution, as well as to external researchers in collaboration. The goal was to increase and take advantage of data and new technologies to better understand teachers and learners through Data Science and Artificial Intelligence (AI).

Currently, the Data Hub offers researchers, faculty, and students three main services[2]:

1) **Curation of institutional data collections**. Data collections are provided upon request and are reviewed in collaboration with their data owners to ensure data integrity, confidentiality, and accessibility. The information technology (IT) and data governance departments verify the anonymization and availability of these datasets.
2) **Design and management of data-driven calls.** Data-driven calls are based on open science (van Dijk et al., 2021), and academically driven by areas of Tecnologico de Monterrey that seek to carry out research and development with their data. These calls are open to the global community of practitioners and accompanied by an *academic advisory board* that contributes to the definition of the dataset and the research areas of the call. Furthermore, through these calls, participants are encouraged to publish research articles on Learning Analytics based on the call's dataset, which in turn we document through a data paper.
3) **Promotion of data collections**. The Research Data Hub (datahub.tec.mx) is a platform that stores and makes available open data produced by Tecnologico de Monterrey's researchers. Through a Dataverse platform, researchers can access the documentation of the educational innovation collections and follow the procedure to request the data of interest. Moreover, researchers can deposit and share data from their investigations.

## 3. RESULTS AND FINDINGS

We observed that the Information provided by the Data Hub: 1) reduces the burden for researchers to investigate the correct provenance of data, 2) minimizes the risk of obtaining outdated and previously manipulated data, and 3) provides only data that has undergone thorough procedures to protect privacy (i.e., masking, obfuscation, and removal of sensible variables).

To strengthen the capabilities of the Data Hub, we continually acquire, integrate, and curate data collections to enable research. For this, data across different departments of Tecnologico de Monterrey has been identified and documented to integrate more than 20 datasets related to students, professors, courses, surveys, competencies development, graduates, continuing education, and more. Our datasets contain 27 years of history of students and professors. This includes more than 0.5 million students and 60 thousand professors at high school, undergraduate, and graduate levels, described through more than a thousand variables. In the coming months, the Data Hub

[2] IFE Living Lab & Data Hub. https://ifelldh.tec.mx/en/data-hub

expects to collect more information related to massive online open courses (MOOCs) and textual data from surveys' open questions and student essays codified using Natural Language Processing (NLP).

Data provenance and governance is another key factor in the design of the Data Hub. For every dataset, stakeholders – Data Governance, Legal, IT, and Data Owners – are kept "on the loop" for the decisions made regarding the use of data. Moreover, the Data Hub commits itself to multiple responsibilities in favor of data governance – e.g., logging requests, and implementing protocols to procure safety, anonymity, and data obfuscation. These commitments are documented in a Data Policy that brings confidence to all the stakeholders.

From January 2022 to September 2023, the Data Hub processed 62 data requests from researchers in 18 countries, with 140 researchers participating in research calls and data requests. Research calls are identified as one of the main drivers of requests, making up 80% of these – mainly due to researchers asking for additional data to the one provided in the call's datasets.

The first data-driven call was focused on predicting the early dropout of higher education students, the second call was related to the analysis of student competencies development (based on activities and pieces of evidence), and the third call proposed improving retention strategies and accentuating the social impact of high-performance students selected for participating in a full scholarship program. These data-driven research calls have increased our researchers' collaboration at many levels, including shared efforts with international institutions, professors from several campuses of our institution, and administrative staff of each campus. In this regard, 28% of researchers participating in our calls are from national and international universities, while 45% of the participant teams have inter-institutional collaboration.

Another important strategy that fostered the use of our data is publishing data descriptors. The dataset prepared for the call on student dropout was documented through a data descriptor published in a high-impact journal (Alvarado-Uribe et al., 2022). This descriptor motivated another 25 researchers worldwide to request this curated dataset in one year.

From these data requests, several research products are arising. Currently, seven articles have been published in Scopus-indexed journals and conference proceedings. To keep track of these products we ask researchers to use a special acknowledgement with the identification of the call or the data request. We also keep track of citations to the data descriptor and to the DOIs that identify the curated datasets in the Dataverse platform.

In the first calls, we observed that the researchers that most participated were from the Engineering School, with low expertise in Education but high proficiency in Machine Learning. We think that a matchmaking tool between AI and Education researchers would produce more robust projects.

## 4. LESSONS LEARNED AND IMPLICATIONS FOR THE FUTURE

As discussed by Lukarov et al. (2020) in building a data warehouse for learning analytics, multiple projects have in common the usage of the same learners' data, which require data governance strategies in favor of not investing repetitive efforts around data management. Moreover, these efforts should occur in observance of issues concerning Learning Analytics, like ethics and privacy of the information contained in educational records (Hoel & Chen, 2018). In this matter, the Data Hub served to break down the barriers to accessing institutional data and accelerated the institutional Data Governance initiative.

Our Data Policy was key for Data Owners to trust us as stewards of their data. This policy includes, for instance, validating the credentials of data requesters and reporting potential issues on publications derived from data. Additionally, we can give access to external researchers, but when sensitive data is involved, we request to include one of our professors in the team to avoid data misinterpretations.

Similarly, data acquisition and documentation are found to be not enough to promote its usage: research calls enable the mobilization of data, increasing the reach to more researchers. Additionally, we observed that call datasets serve a wide range of participants since they are co-created in collaboration with primary call stakeholders: advisory board members, call participants, Data Owners, and the Data Hub team. As evidence, the publication of the data descriptor led to an extended reach of the call, having 25 requests of the dropout dataset in one year - more than double the number of teams that participated in the corresponding call.

With more than 1,000 variables from more than 20 data collections, finding the variables of interest becomes a challenge. As a solution, a naming convention is being developed to identify variables and ease joining data from different collections. Data request processes are being reviewed and automated to facilitate researchers to identify the data (population, variables, and filters) they need.

Another challenge we face now is deciding between using our resources to increase our datasets, clean and document data, and generate new data-driven research calls. We are testing short programs such as workshops where researchers working on a specific topic dialogue with data owners to match needs and expertise, while we use the already integrated data.

We expect to continue documenting the data usage and provide a formal evaluation of the efficiency of the data distribution channels: dataset in the data portal, data-driven call, and data descriptors. We are also surveying our stakeholders to identify the main (dis)advantages of our initiative.

## 5. REFERENCES

Alvarado-Uribe, J., Mejía-Almada, P., Masetto Herrera, A. L., Molontay, R., Hilliger, I., Hegde, V., Montemayor Gallegos, J. E., Ramírez Díaz, R. A., & Ceballos, H. G. (2022). Student Dataset from Tecnologico de Monterrey in Mexico to Predict Dropout in Higher Education. *Data*, *7*(9), 119. https://doi.org/10.3390/data7090119

Hoel, T., & Chen, W. (2018). Advancing the delicate issue of ethics and privacy for Learning Analytics. *Companion Proceedings 10th International Conference on Learning Analytics & Knowledge (LAK20)*, 149–150.

Lukarov, V., Ehlenz, M., & Schroeder, U. (2020). Building a data warehouse for multimodal learning analytics research projects. *Companion Proceedings 10th International Conference on Learning Analytics & Knowledge (LAK20)*, 25–28.

van Dijk, W., Schatschneider, C., & Hart, S. (2021). Open Science in Education Sciences. Journal of Learning Disabilities, 54(2), pp 139-152.

# Platform-based Adaptive Experimental Research in Education: Lessons Learned from Digital Learning Challenge

**Ilya Musabirov[1], Mohi Reza[1], Steven Moore[2], Pan Chen[1], Harsh Kumar[1], Tong Li[1], Haochen Song[1], Jiakai Shi[1], Koby Choy[1], Thomas Price[3], John Stamper[2], Norman Bier[2], Nina Deliu[4], Sofia Villar[5], Anna Rafferty[6], Audrey Durand[7], Joseph Jay Williams[1]**

[1] University of Toronto, [2] Carnegie Mellon University, [3] North Carolina State University, [4] Sapienza Università di Roma, [5] University of Cambridge, [6] Carleton College, [7] Université Laval

**ABSTRACT**: We report on our experience with a real-world, multi-experimental evaluation of an adaptive experimentation platform within the XPRIZE Digital Learning Challenge framework. We showcase how EASI (Experiment as a Service) cross-platform software supports quick integration and deployment of adaptive experiments as well as five systematic replications within a 30-day timeframe. The outline the key scenarios of the applicability of platform-supported experiments and reflect on lessons learned from this two-year project that can help researchers and practitioners to integrate adaptive experiments in real-world courses.

**Keywords**: adaptive experiments, posterior sampling, experimentation platforms

## 1 INTRODUCTION

In this paper, we report on our experience with a multi-experiment field deployment for evaluating an adaptive experimentation platform within the XPRIZE Digital Learning Challenge (DLC) framework, which took place between March 2022 and 2023. We show how we used our Experiment as a Service (EASI) cross-platform software infrastructure for experimentation to conduct and systematically replicate five experiments within 30 days. EASI has been used for a diverse range of over 250 traditional experiments to date. Unique strengths include a range of existing random assignment methods and usage across many different settings, whereas most tools are overly specific to the context/platform of creation. Designed for interoperability, EASI has been used in different digital learning platforms (edX, Coursera, Moodle, Canvas, ASSISTments), and can be integrated with any LTI (Learning Tools Interoperability) compliant LMS. EASI provides access to a library of machine learning algorithms and statistical methods (such as Bayesian inference) for analyzing data in real time. This approach offers flexibility in changing how conditions are assigned to future students and how instructors and students can be involved in classroom experimentation (Reza et al., 2021). We report the key lessons on platform-supported adaptive experiments learned from our work in the XPRIZE DLC[1].

[1] https://www.xprize.org/challenge/digitallearning

## 2 ADAPTING EXPERIMENTS PROPORTIONAL TO UNCERTAINTY

Instructors have reasonable concerns about experiments being fair and assigning students to a worse condition. When one experiment provides some evidence for a difference in conditions, instructors may be even more reluctant to do replications to better understand the design choices that led to an intervention's effectiveness or how underrepresented learner groups benefited. However, it is practically and scientifically essential to become increasingly certain that our intervention is effective, or test out potentially better ideas. We believe the apparent dichotomy—to experiment or not—is better framed as a fundamental statistics and Machine Learning (ML) tradeoff between 'exploring' (collecting data to converge on the best action) and 'exploiting' (exploiting/using the data to change which actions we choose).

Traditional experiments assign conditions uniformly, while adaptive experiments adjust assignment probabilities based on the latest evidence. EASI turns any single deployment into a rapid sequence of replications since data from even the past one, five, or ten students can gradually adjust the design of the experiment. EASI uses a Posterior Sampling Algorithm (Chapelle & Li, 2011) to analyze individual participant data, updating the likelihood of assigning conditions based on student data collected from a course. When dealing with distinct student groups, Contextual Posterior Sampling is used to determine the optimal condition for each group of students. Over time, as more student data accumulates, the approach shifts from a standard experimental split (i.e. 50/50) toward a more personalized one for each student group. This algorithm enables customized interventions for each group, enhancing their learning results.

## 3 FIELD DEPLOYMENT: XPRIZE DIGITAL LEARNING CHALLENGE (DLC)

During the IES-sponsored DLC, we aimed to demonstrate the adaptive approach to experimentation and our platform capabilities in the rapid multi-replication of educational interventions for different student demographics. To achieve this, we integrated EASI with the Open Learning Initiative (OLI) (Bier et al., 2023), which is designed to support robust experimentation at scale, in collaboration with institutions that are already using OLI courseware. The institutions include R1 universities and community colleges. Instructors of these courses had previously used the OLI platform for one to three semesters prior, without EASI integrated into it. We reached out to them for their consent and, with IRB approval, integrated EASI into these courses. We performed one pilot and five replication studies in distinct courses. These courses are offered across five diverse institutions, serving 2295 enrolled students in several domains, from Anatomy to Statistics.

Our iterative design approach involved an interdisciplinary working group of EASI and OLI developers, machine learning, learning science, and engineering researchers in collaboration on all stages of the DLC, focusing on eliciting critical scenarios of using adaptive experimentation for continuous improvement in the set of diverse courses. Three team members were also course instructors with extensive experience using OLI, while another two had expertise in other LMS.

Tu support deployment in a diverse set of courses, we chose intervention cases which, with the help of EASI and OLI, were designed as loosely coupled with the course content and easily portable, allowing us as well as researchers and instructors to rapidly replicate them in any new course, using existing course content. These interventions were designed for a common formal education context where students independently work through an online textbook containing short passages and videos with content knowledge. At the end of each section of the textbook, students engage in a variety of activities to promote learning. We aimed the first intervention case at the motivational domain, encouraging students to participate in optional course activities, randomizing Growth Mindset, and Self/Peer-Focused Framing, and using engagement outcomes. The second intervention aimed to provide students with retrieval practice prompts tied to course activities and used accuracy on the following problem as a proximal outcome (algorithm reward). We used these designs to discuss how adaptive interventions can help to explore and replicate the impact of novel variations of existing interventions, precisely targeting various outcomes of interest (e.g., assessments and participation in learning activities).

## 4 LESSONS LEARNED

In the iterative design of experiment and replications, we elicited **key adaptive scenarios** that can accelerate analysis and action using Bayesian statistics and ML algorithms: **(S1)** dynamically adding a new condition to expand from a two- into a three-condition experiment; **(S2)** how a three-condition study can assign more students to the most effective conditions while increasing statistical power to identify which ones are the best; and **(S3)** balancing practical impact with scientific insight, by helping a majority group get conditions that are better on average (for that group), while still collecting data to personalize so that a statistical minority is not unfairly receiving a condition that is worse for them. These features help researchers and instructors interested in course improvement to conduct a broad range of adaptive experiments. They can target micro-level objectives, such as enabling specific actions or improving a particular course element, or macro-level goals, such as improving overall course outcomes. Moreover, the dynamic assignment of students to better conditions decreases the decision-making burden on the course team, while allowing for improved student outcomes. It also has the potential to increase statistical power while better discriminating between conditions in multi-armed cases.

Another critical focus should be showing the consequences of particular adaptive experimental patterns in the early design stages using **statistical simulations**. During the DLC, we prototyped modules aimed to let instructors/researchers take data from one experiment and use it to specify alternative Scenarios for what the effects of conditions might be in different replications. That helps researchers simulate collecting and analyzing data from thousands of repeated runs of an experiment under different scenarios for what the effects could be and what might mediate these effects, such as student characteristics. This allows researchers and instructors to specify precisely and explore different kinds of effects they could discover in future replications of their experiment and understand what impact the particular adaptive experimental design can achieve compared to traditional approaches. Another related requirement is providing a set of custom **data visualizations and data analysis workflows** tied to the experimental design. This avoids potential issues arising from the application of unsuitable or suboptimal analytical methods. It allows us to understand not only what we have learned – causal effects, but also how we did – the impact of adaptation and personalization on students.

We illustrate these two connected tools in the simulated example, based on **S3**, with two groups of students showing higher and lower accuracy. In this scenario, students are encouraged to contribute their own questions to the course bank using two approaches: Self-Focused condition (focused on utility for student's learning) and Peer-Focused condition (encouraging sharing knowledge with others). Overall, it looks like the Peer-Focused condition is better than the Self-Focused condition ($z= 4.58$. $p<0.0001$, Cohen's w 0.12). However, a closer analysis would reveal that for a statistical minority (20% of students with Lower Accuracy), the opposite is true: the Peer-Focused condition is worse than the Self-Focused condition, with a larger effect size ($z= 3.52$. $p=0.0004$, Cohen's w 0.2). The Self-Focused message being better for Lower Accuracy students is obscured by the fact that 80% of the students are in the Higher Accuracy group, where the opposite is true: The Peer-Focused condition is better than the Self-Focused condition ($z=6.91$. $p<0.0001$, Cohen's w 0.2). This effect in the 80% Higher Accuracy students drives the overall average positive effect, although giving everyone the Peer-Focused prompts is harmful to Lower Accuracy students. A suitable adaptive intervention template can automatically account for this crossover interaction. In many similar scenarios the potential of adaptivity and/or personalization needs to be well-communicated, and experimental platforms need to support their users in exploring and visualizing them early on to make informed decisions about intervention designs, especially from the equity perspective.

The last direction s providing researchers and course teams with **pre-defined templates for adaptive experimentation**, capturing potential decision points to intervene in the course, meaningful outcomes to use in the adaptive experiment, a template and examples for the content part of the intervention, and potential bandit designs (e.g. three-arm non-contextual, contextual based on previous performance) with their impact.

## 5 CONCLUSION

Our work during the DLC has emphasized the challenges of experimentation in post-secondary settings, particularly as educators adapt and modify their planned instructional activities. The ability to flexibly adjust to these types of changes is an essential characteristic of any educational experimentation platform. We highlight the advantages of adaptive experiments: the ability to adjust experiments on the fly, based on real-time data, can lead to more efficient and effective research outcomes, ultimately helping to accelerate progress in the field of education. By leveraging platforms like EASI, researchers, and educators can gain new insights into the most effective teaching and learning strategies, paving the way for improved student outcomes and a brighter future for all. This work was partially supported by the National Science Foundation (#2209819).

## REFERENCES


Bier, N., Stamper, J. , Moore S., Siegel, D., & Anbar, A. (2023). OLI Torus: a next-generation, open platform for adaptive courseware development, delivery, and research. Companion Proceedings of the 13th International Learning Analytics and Knowledge Conference.

Chapelle, O., & Li, L. (2011). An Empirical Evaluation of Thompson Sampling. Advances in Neural Information Processing Systems, Vol. 24. Curran Associates, Inc.

Koedinger, K. R., Stamper, J. C., Leber, B., & Skogsholm, A. (2013). LearnLab's DataShop: A Data Repository and Analytics Tool Set for Cognitive Science. Top. Cogn. Sci., 5(3), 668-669.

Reza, M., Kim, J., Bhattacharjee, A., Rafferty, A., & Williams, J. (2021). The MOOClet Framework: Unifying Experimentation, Dynamic Improvement, and Personalization in Online Courses. Proceedings of the Eighth ACM Conference on Learning @ Scale.

# Hackathons for Awareness and Community Engagement in Learning Analytics

**Craig Thompson, Alison Myers, Justin Lee, Will Engle**
The University of British Columbia
craig.thompson@ubc.ca, alison.myers@ubc.ca, justin.lee@ubc.ca, will.engle@ubc.ca

**ABSTRACT**: **Practitioner Presentation.** Challenges of institutional adoption of learning analytics include lack of student engagement, little transparency, and few opportunities for feedback. In this report, we reflect on one element of our institution's approach to these challenges: regular engagement of students, practitioners, and institutional leadership through Learning Analytics Hackathons. The history and evolution of these hackathons mirror the advancement of learning analytics at the University of British Columbia: they started as community-driven events but have gained institutional support and attention as our data infrastructure and learning analytics culture matures. At the time of writing, the authors are planning the 9th Learning Analytics Hackathon[1]. In this report, we share our approach, lessons learned, and discuss opportunities for continued student engagement in learning analytics at our institution.

**Keywords**: Hackathon, Student Engagement, Institutional Learning Analytics, Stakeholders

## 1 INTRODUCTION

Hackathons are intensive events that bring participants with a wide range of skills and backgrounds together to both collaborate and compete to solve a problem or create something innovative within a short timeframe. Students participate in hackathons for various reasons, including the prospect of social interaction and fun, the opportunity to learn new skills and knowledge, the possibility to win prizes, and the chance to apply classroom knowledge in real-life scenarios (Nandi & Mandernach, 2016; Steglich et al, 2021). University-based hackathons often sit in a unique space outside of formal curricula or program requirements. Instead, students "donate" their time, often giving up several days in a row, to attend such events.

The use of hackathons as a strategy for student engagement by institutional learning analytics initiatives represents an opportunity to promote learning analytics as a field, to teach analytics skills necessary for researchers and practitioners, and to educate and engage students in learning analytics principles and practices. The past decade has seen learning analytics education occurring in a variety of formats including graduate degrees and courses, undergraduate certificates, MOOCs (Kizilcec & Davis, 2023), and events such as LAKathons[2] (which have been a regular occurrence at LAK since 2015) and workshops at LASI[3]. Hackathons are an opportunity to engage in student-centered analytics by involving students in the design of learning analytics tools, moving away from a "black box" of analytics and towards a paradigm of "Glassbox Analytics" (Ochoa & Wise, 2021) by creating transparency of data collection and use at the institution. Finally, these events may help students interpret analytics

[1] https://github.com/UBC-LA-Hackathon/hack-la-2023
[2] https://lakathon.org/history/
[3] https://www.solaresearch.org/events/lasi/

and frame it not just as a tool or data, but as information that can be used to support their learning by introducing them to the field and techniques of learning analytics.

## 2 APPROACH

Hackathons at the University of British Columbia (UBC) take place over a weekend and are held in learning spaces designed for collaboration. The introduction to the event is an overview of the data (where it comes from, possible challenges of interpretation, and whether the data is representative of data collected at UBC) and the challenge we are asking students to tackle. Any important information about data privacy, ethics, and security is reviewed. These events offer opportunities for students to learn about the kinds of data being collected for the purposes of learning analytics – as much as possible we aim to share data with students that gives them a realistic view into these kinds of data, in particular, what is collected or used at UBC. Students are then free to "hack"; they work in teams with the data provided to solve a challenge presented to them or develop their own learning analytics tools which they will present in a lightning talk. Expert volunteers from the learning analytics community at UBC are available to provide support throughout. Volunteers have included programmers, experts in Open Education, data analysts, and LA researchers and practitioners (to name a few).

At the end of the event, students are invited to present their work in a lightning talk. Teams are provided with brief feedback and, depending on the specific event, top projects are recognized. This evaluation is part of an organizing strategy to blend light competition with a strong undercurrent of collaboration. During the short timeframe of the hackathon, participants within and across teams share ideas, solve problems, and lend expertise to fellow participants. This collaborative spirit not only fosters a sense of community but also leads to the creation of more robust and imaginative projects. Winning teams receive small prizes (e.g., gift cards to the University bookstore) and additional feedback from the hackathon organizers. Breakfast and lunch are provided throughout the hackathon and all attendees earn certificates of participation.

One of our goals is to engage with a diverse range of students with various levels of technical experience and skills; thus, in addition to our traditional hackathons, we alternate with events consisting of a series of workshops that introduce skills or techniques relevant to learning analytics. These "workshop-a-thons" are organized along similar principles and timeframes as the hackathons but involve more structure and mentor facilitation. Workshops are coordinated so that students are scaffolded through designing a tool or solution. Workshop topics range from ideation/design thinking for learning analytics to more technical sessions on data wrangling, data visualization, and software development. Students attend a workshop, have "hack" time where they apply what they learned, and then attend the next workshop in the series. For examples of the challenges presented and outcomes[4] from both event types see Table 1.

## 3 OUTCOMES

Since 2015, over 600 students have participated in our learning analytic hackathons; the events are popular and registration (which is capped due to room size) is always at capacity. The hackathons have promoted robust student engagement by immersing them in a dynamic, problem-solving environment where they actively contribute to solving real-world educational challenges. This hands-

[4] The events are typically covered by internal reporting at our institution and summaries can be found at https://learninganalytics.ubc.ca/for-students/hackathons/

on engagement also empowers students with practical skills. Moreover, the hackathons serve as a platform for increasing student awareness of learning analytics.

**Table 1. Hackathon and Workshop-a-thon Examples**

| *Event Type* | *Examples of Past Events* | *Student Projects* |
|---|---|---|
| Hackathon (80-100 students + mentors) | Use the Canvas REST API to extract your own data and develop a tool, analysis, or app that you would find useful for your own learning.<br>Using deidentified data from Canvas and other supporting course information, what tools would you develop to support an instructor? | Gamified learning in Canvas – where you earn points for certain interactions; Grade predictions using a neural net model; Quiz and assignment statistic visualizations. |
| Workshop-a-thon (50-60 students + mentors) | Using data from an EdX MOOC, take the role of an LA expert: what would you do with this data to improve teaching or learning? You have the option of working with a team, individually, or attending our series of workshops to get a better understanding of some of the techniques that may be useful (R for data analysis, Tableau for visualization). | Interactive tree/Sankey diagrams showing student interaction with course content; Visualizations showing play/pause/seek behaviour of video watching. |

Hackathons have helped UBC's learning analytics staff gain insights into student interests both through the tools that students develop and conversations with students throughout the event. Additionally, the hackathons have helped build capacity within our institution for creating open resources including deidentified and cleaned data, allowing for safer and more responsible data sharing, and fostering a culture of transparency and collaboration. We have been able to use "real" data which has been scrubbed of any identifiable information through a combination of hashing data, noising data (I.e. adjustment of grades +/- certain points, changing timestamps to the future), completely removing any personally identifying data, and masking certain information (Khalil and Ebner, 2016). In addition, the event planning process often involves the development of policies and procedures for students to access data, ensuring that sensitive information is handled with care and in compliance with university policy and privacy regulations.

The organization of the hackathons has also been an opportunity for internal event promotion and summaries which include excerpts from interviews with collaborators and students. Students are enthusiastic about the experiential learning and the opportunity to contribute to the university: "Contributions to Canvas, even minor developments, can benefit tens of thousands of [masked] students by either improving efficiencies or providing insights," said one student, continuing, "I regarded this hackathon as something I could do for our community" (2018). Collaborators recognized a lack of student voice and perspective in the implementation of learning analytics: "At the moment, most universities' learning analytics designs are for instructors and researchers. We don't see a lot of designs for students" (2018). Students frequently attend without expecting substantial prizes or incentives, driven solely by their genuine curiosity: "I came to the Hackathon because I was really curious about what could be done with the learning analytics side of data analysis," said one student (2019)**.** "I just wanted to get my feet wet and see what I could come up with."

## 4 LOOKING FORWARD

The hackathons became recurring institutionally supported events during UBC's "pilot" phase of learning analytics. In this phase, engagement with stakeholders and feedback that might inform our institutional approach to learning analytics was a priority. One goal of the organizers was for hackathons to act as pipelines for student work to possibly become pilot initiatives – where we could offer continued opportunities for students to continue to build learning analytics tools alongside the learning analytics team. However, this has been a significant challenge to achieve, primarily due to lack of appropriate data infrastructure, budget for student employees, and a fair method to select which projects might enter the pipeline. The authors hope that with continued improvement to our data infrastructure this may still one day be an achievable goal. Regardless, we have found that hackathons can play a crucial role in engaging students in learning analytics institutionally. They provide a platform for students to apply their skills, collaborate, and gain practical knowledge while also contributing to the development of learning analytics tools and practices. A combination of low-barrier entry with the inclusion of workshop-a-thons, and a team of volunteers allows students at all levels to participate. These events can promote community engagement and help institutions better understand and involve students in shaping the future of learning analytics.

## 5 REFERENCES

Khalil, M., & Ebner, M. (2016). De-Identification in Learning Analytics. *Journal of Learning Analytics*, *3*(1), Article 1. https://doi.org/10.18608/jla.2016.31.8

Kizilcec, R. F., & Davis, D. (2023). Learning Analytics Education: A Case Study, Review of Current Programs, and Recommendations for Instructors. In O. Viberg & Å. Grönlund (Eds.), *Practicable Learning Analytics* (pp. 133–154). Springer International Publishing. https://doi.org/10.1007/978-3-031-27646-0_8

Nandi, A., & Mandernach, M. (2016). Hackathons as an Informal Learning Platform. *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, 346–351. https://doi.org/10.1145/2839509.2844590

Ochoa, X., & Wise, A. F. (2021). Supporting the shift to digital with student-centered learning analytics. *Educational Technology Research and Development*, *69*(1), 357–361. https://doi.org/10.1007/s11423-020-09882-2

Steglich, C., Marczak, S., Guerra, L., Trindade, C., Dutra, A., & Bacelo, A. (2021). An online educational hackathon to foster professional skills and intense collaboration on software engineering students. *Brazilian Symposium on Software Engineering* (pp. 388–397). Association for Computing Machinery (ACM). https://doi.org/10.1145/3474624.3476973

# Implementing Learning Analytics in Data-poor Contexts: The Role of an Institutional Data Audit

**Christine I. Ofulue, Adewale Adesina, Felix K. Olakulehin, Oluwaseun P. Oluyide**
National Open University of Nigeria
[cofulue, aadesina, folakulehin, ooluyide] @noun.edu.ng

**ABSTRACT**: The advent of Learning Analytics offers educational institutions the opportunity to enhance their educational outcomes through data-driven insights. However, the effective deployment of Learning Analytics relies on an understanding of an institution's existing data landscape. This case study outlines a data audit approach for implementing Learning Analytics at the National Open University of Nigeria (NOUN), West Africa's largest distance education provider. It highlights essential considerations for deploying such systems in contexts often characterised as data-poor, where there is a shortage of digital data and infrastructure and institutional capacity for strategic utilisation of learning data. An in-depth data audit was conducted at the NOUN using focus group interviews with key stakeholders. The audit sought to identify existing data, data ownership, and data release protocols. The findings from this audit unveiled a wealth of untapped data and highlighted specific areas demanding strategic attention to maximise data utilisation. This case study serves as a valuable roadmap for educational institutions aspiring to harness the potential of Learning Analytics while grappling with the complexities of their existing data ecosystems. It underscores the pivotal role of institutional data audits in establishing a foundation for informed, effective, and ethically sound Learning Analytics practices.

**Keywords**: data audit, data-poor, learning analytics, stakeholders, data utilisation

## 1 INTRODUCTION

Learning analytics (LA) provides the opportunity to collect, analyse and use students' data for understanding and optimising learning and the environments in which it occurs (Long, Siemens, Conole & Gašević, 2011). Central to the implementation of LA is student data (Knight, Gibson & Shibani, 2020). However, in data-poor contexts- whether referring to a lack of (digital) data and digital infrastructures, or the lack of institutional capacity to strategically utilise, analyse and operationalise learning data, student data is either not available or accessible for operationalising LA. This paper presents a case study of a data audit at National Open University of Nigeria (NOUN) as part of the implementation of LA. This paper reports the rationale, approach and outcomes of a qualitative data audit process adopted by the LA research and implementation team at NOUN.

## 2 BACKGROUND AND CONTEXT

The NOUN, the first public open and distance learning university in West Africa, has over 120,000 active students. Established to increase higher education access and bridge enrolment gaps in Nigeria, it has achieved diverse student enrolment across Nigeria and some foreign countries since 2002. While NOUN appears successful in increasing university education enrolment, it faces challenges, notably a high student attrition rate. For example, of over half a million enrolled students, only around 120,000 consistently sit for end-of-semester exams over the past six years (Institution's MIS, Interview 2023). Student dropout has been an "evergreen" problem throughout the evolution of distance education (Elibol & Bozkurt, 2023) and a source of significant concern for distance education providers, including NOUN.

To address the issue, the institution adopted LA as a tool to understand student success, and risk potentials in mitigating attrition and drop-out (Guzmán-Valenzuela et al., 2021). A dedicated LA research team was formed to develop an institutional framework, aimed at enhancing student success through improving learning design and student support, using the SHEILA framework. Recognising the need for a comprehensive understanding of available data and personnel capacity for LA, the team initiated a data audit at NOUN. The audit process is discussed in the following section.

# 3 DESIGNING THE DATA AUDIT

Research on Learning Analytics (LA) often neglects the importance of understanding institutional data access, hosting, and governance. Despite its significance, frameworks like SHEILA omit initial data audits (Tsai et al., 2018). Audits assess data quality and utility, offering efficiency and risk management benefits in LA (Jones et al., 2008). Success depends on data quality, institutional awareness, and ownership identification, crucial for resource prioritisation and risk mitigation. Effective LA relies on meaningful data access and utilisation for pedagogical purposes (Schläppy, 2016).

Recognising the potential of Learning Analytics (LA) to address student concerns at NOUN, the project team embarked on operationalising LA by identifying existing information sources across the university. This includes data from prospective learners, students, former learners, and staff, though the nature, format, and sources of the data remained unclear. To advance the LA project, the team prioritised determining data ownership, usage, and assessing current skills and capacities at NOUN. Conducting a thorough data audit was the initial step, evaluating available data, identifying gaps, overlaps, and missing aspects. This process is crucial as it informs the development of tailored policies, practices, and processes to meet institutional and student needs. NOUN's diverse data-generating units, including Examinations and Assessment, Management Information Systems, Learner Support Systems, among others, play pivotal roles in this endeavour.

# 4 DOING THE DATA AUDIT

The methodology adopted to establish the data-scape at NOUN was in the form of focus-group structured interviews with the heads/key representatives of each of the departments identified above. Invitations were sent to the heads of departments, and they were informed of the purpose of the interview which was to get a sense of their data collection process, the available data, and how such data are stored. The interviews lasted between 45 minutes and 1.5 hours and were held over one week. The key informants were asked the same questions, and prompts and clarifications were provided where they were required by the participants.

## 4.1 Focus Group Discussion (FGD) Interview schedule

The project aimed to understand the profiles of successful and at-risk students and improve the learning environment accordingly. This required investigating available data on students' learning experiences across the university. Focus group discussions (FGDs) with department heads/key representatives were structured with ten linked questions, ensuring a logical flow of information from follow-up questions.

1. What data do you have in your department?
2. In what format does the data exist: analogue or digital?
3. Did your department originally collect it?
4. If NO to 3 above, do you have access to this data collected by another department?
5. Who hosts the data? And where are they hosted e.g., server, cloud, hard drives?
6. How is the data currently used?
7. If you host the data, who has access to it?
8. What levels of access currently exist?
9. If you host the data, who grants access?
10. Under what conditions can the data be shared?

Responses to the above questions by each of the data-generating departments of the university were collected in tabular format. These are accessible in an online drive at Data Audit Survey

Additionally, detailed responses received to the questions which came from prompts and extended conversations were recorded by the researchers in their field notes. After each interview, we compared notes to agree on perspectives and responses to issues and these are categorised into themes. At the end of the interview exercise, we also compared responses across each of the departmental responses to deepen our understanding of how available data might fit into the overall LA process of the university.

# 5 THEMATIC RESULTS FROM THE DATA AUDIT

From the data collected some of the major themes that emerged from the audit are as follows:

A. There is a **lot of data** (automated, directed and gifted) - scattered across the institution - some analogue but most data are digital/digitised.
B. There is currently **no seamless access to, and/or integration** of stored data
C. There is a need for a strategic and institution-wide analysis of **current and future data needs** to inform **pedagogy and support**.
D. Students are currently mainly **seen as data producers** and not **data users**.
E. The different departments **do not know what data are available** from the different units and therefore **duplicate** data collection.
F. Data collected in one section (e.g., application or quality assurance unit) is **not pushed or made available** further in students’ learning journey (e.g., course registration).
G. **The value of data** is not fully realised.
H. Push and pull on-demand - **automated collection, no automated push or use**.
I. There is a **need for robust data governance**.
J. **The democratisation of data** - seamless - everyone with a data need will have appropriate access.

The themes that emerge following an interrogation of the data obtained from the data-generating departments were used to frame the LA process for NOUN.

## 6 PRACTICAL LESSONS FROM THE DATA AUDIT

Practical lessons from the data audit underscored the importance of the quality of data collected for informed decisions, ethical use ensuring student privacy, and governance for data protection, management, and utilisation in refining learning analytics strategies.

## 7 ACTION STEPS BASED ON FINDINGS FROM THE DATA AUDIT

The next steps for Learning Analytics (LA) at NOUN involve leveraging on the insights from the data audit, which revealed an abundance of underutilised data. Establishing a central repository and implementing a robust data governance framework are priorities to address accessibility and integration challenges across departments. An institution-wide skills/capacity audit will inform pedagogy and support strategies. Six specific action-steps include sensitisation workshops, piloting studies, centralising data management, creating user-friendly interfaces, enforcing data governance protocols, and staff training on analytical techniques. These measures aim to optimise data potential, improve learning experiences, and tackle student attrition. They lay the groundwork for ethical LA practices, enhancing educational outcomes, and mitigating educational wastage, ensuring NOUN's commitment to quality pedagogy and student success.

## 8 CONCLUSION

In conclusion, understanding existing data holdings is crucial for resource prioritisation and risk management in learning analytics. Data audits ensure accuracy, integrity, and ethical use, vital for institutions like NOUN that are data utilisation poor. Feedback refines strategies, addressing challenges of accessibility and utilization, ultimately enhancing data quality and decision-making.

## REFERENCES


Guzmán-Valenzuela, C., Gómez-González, C., Rojas-Murphy Tagle, A., & Lorca-Vyhmeister, A. (2021). Learning analytics in higher education: a preponderance of analytics but very little learning? International Journal of Educational Technology in Higher Education, 18, 1-19.

Jones, S., Ball, A., & Ekmekcioglu, Ç. (2008). The data audit framework: a first step in the data management challenge. The International Journal of Digital Curation, 3(2), 112-120

Knight, S., Gibson, A., & Shibani, A. (2020). Implementing learning analytics for learning impact: Taking tools to task. The Internet and Higher Education, 45, 100729.

Long, P. D., Siemens, G., Conole, G., & Gašević, D. (Eds.). (2011). In Proceedings of the 1st International Conference on Learning Analytics and Knowledge (LAK'11). Retrieved from https://tekri.athabascau.ca/analytics/

Schläppy, M. L. (2016). Learning analytics at the University of the Highlands and Islands.

Tsai, Y. S., Rates, D., Moreno-Marcos, P. M., Muñoz-Merino, P. J., Jivet, I., Scheffel, M., ... & Gašević, D. (2020). Learning analytics in European higher education—Trends and barriers. Computers & Education, 155, 103933.

# Predicting Problem-Solving Success in an Office Simulation Using N-Grams and Random Forest on Behavioral Process Data

**Sabrina Ludwig**
Mannheim Business School, University of Mannheim, Mannheim, Germany
Sabrina.ludwig@uni-mannheim.de

**Andreas Rausch**
Mannheim Business School, University of Mannheim, Mannheim, Germany
rausch@uni-mannheim.de

**ABSTRACT**: Early prediction of students' problem-solving success in computer-based simulations enables personalized support from teachers or adaptive educational systems. This paper employs the Random Forest machine learning model to predict students' problem-solving success in a 55-minute business-related scenario. Early behavioral data (within the first five, ten, and 20 minutes) based pm approximately 23,000 mouse clicks and keyboard strokes were recorded for 234 trainees, reflecting their problem-solving behavior. Utilizing n-gram sequence mining, a technique widely recognized in natural language processing and machine learning, we trained the model on both all features (2-grams) and selected features with high predictability. Results indicate that accurate predictions were possible after the first ten and twenty minutes, but not after only five minutes. As the early-window size increased, classification performance improved. The model using selected features from the first 20 minutes achieved the highest AUC score of approximately .70. This accuracy level aligns with similar studies. These predictions offer instructors a valuable tool for identifying struggling students early and providing tailored support, while also allowing for adaptive enrichment of tasks for more successful students.

**Keywords**: Prediction, Problem-solving, Log data, n-gram, Random Forest, Simulations

## 1. INTRODUCTION

Log data is valuable for understanding student problem-solving behavior in computer-based learning environments. Educational data mining (EDM) enables the analysis of problem-solving behavior by extracting cognitive processes from recorded log data. Mining sequence patterns, particularly n-grams, divide complete action sequences into smaller units (Han et al., 2019). Taking into account the frequency of multiple actions (bi- or trigrams) provides greater insight compared to uni-grams (He & von Davier, 2015), as it examines the movement from one activity to another (such as writing notes following reading a document). Machine learning models like Random Forest help to identify struggling students early (Tomasevic et al., 2019). Analyzing problem-solving processes enables early predictions of success, fostering learning through prompts (Li et al., 2017; Lu et al., 2018). Several studies exist that focus on n-grams and classification models for simulating behavior. For instance, Brandl et al. (2021) proposed random forest to predict medical students' diagnostic accuracy based on collaborative diagnostic activities (bi-grams). Another study utilized XGBoost and log data to forecast success and failure in early stages of the 'Programme for the International Assessment of Adult Competencies (PIAAC)' (Utilizsch et al., 2022). Various studies suggest that machine learning models hold promise in predicting problem-solving success. However, due to the unique nature of simulation environments, the generalizability of these findings for business-related office simulations remains unclear. Therefore, we address two research questions:

(1) How do groups with different levels of problem-solving success differ in their problem-solving behavior in a computer-based office simulation?
(2) How early can problem-solving success be predicted based on problem-solving behavior?

## 2. Method

To answer these questions, we used a comprehensive office simulation with typical office tools and several complex problem-oriented scenarios. We collected protocol data from 234 vocational students to investigate early problem-solving behavior in a computer-based office simulation (Figure 1). In the office simulation, the students first familiarized themselves with the office tools during an onboarding tutorial. They then participated in a dynamic supplier selection scenario. The scenario began with an email involving the task assignment. Learners examined both relevant and irrelevant documents and conducted a cost-benefit analysis using a spreadsheet, before concluding with a supplier decision communicated via response email. The performance evaluation included a comprehensive coding system that incorporated the cost-benefit analysis and the response email. Regarding the analysis, participants were divided into two groups based on their performance scores in binary classification and prediction. N-grams derived from early behavior served as an input to the classification model. Random forest was used to classify participants based on the early behavior data.

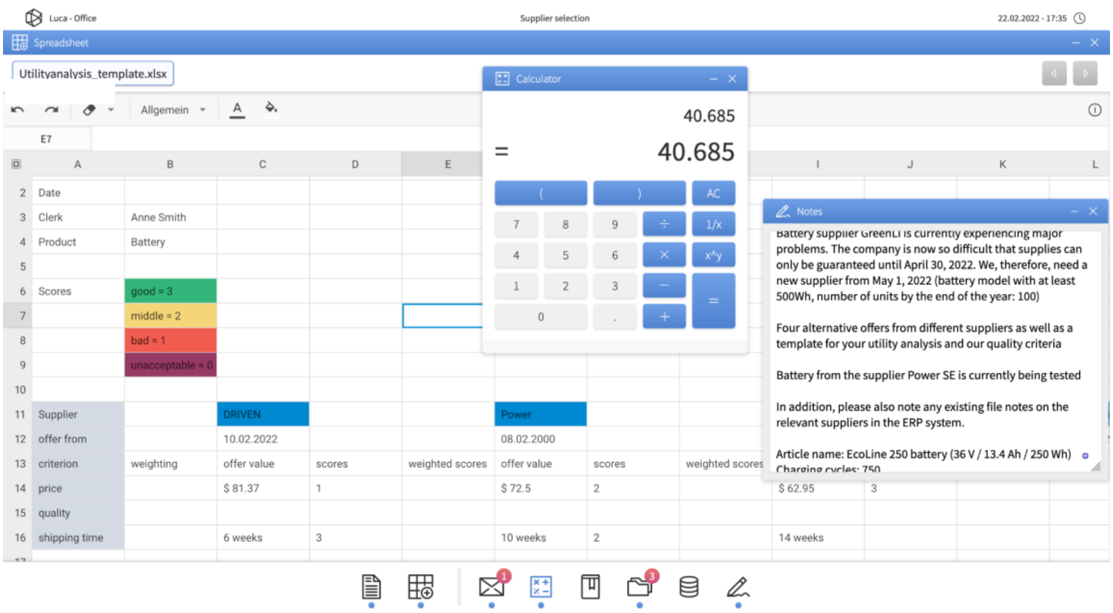



**Figure 2:** Screenshot of the spreadsheet, notepad and calculator in the LUCA Office Simulation

## 3. Findings

Descriptives highlight similarities and differences between more and less successful participants. Higher problem-solving success corresponds to greater activity frequency in the first five minutes (4.245 vs. 4.205), ten minutes (7.414 vs. 6.706), and twenty minutes (15.264 vs. 13.850). Evaluation scores were computed for each model (see Table 1). M1 exhibited poor accuracy at .53, and a low AUC ROC score of .61 for classifying based on the first five minutes. As the early-window size increased, accuracy improved to .60 for both M2 (first 10 minutes) and M3 (first 20 minutes). Additionally, the AUC score saw enhancements, reaching .63 and .67 for M2 and M3, respectively, using features from the first 10 and 20 minutes. We performed feature selection to enhance the classification AUC Score, considering potential redundancies in the datasets. Figure 2 illustrates the impact of feature selection on model performance across early time intervals. Selecting robust features within the first 5, 10, and 20 minutes notably improved AUC scores (from .61 to .65 for $M1_{selected}$, .63 to .66 for $M2_{selected}$, and almost .70 for $M3_{selected}$). Notably, M6's score aligns with findings in similar studies, as discussed below.

**Table 1** Evaluation of model's performance (without feature selection)

| | Random forest model (M1: first 5 min) | Random forest model 2 (M2: first 10 min) | Random forest model 3 (M3: first 20 min) |
|---|---|---|---|
| Accuracy | .53 | .60 | .60 |
| AUC ROC | .61 | .63 | .67 |

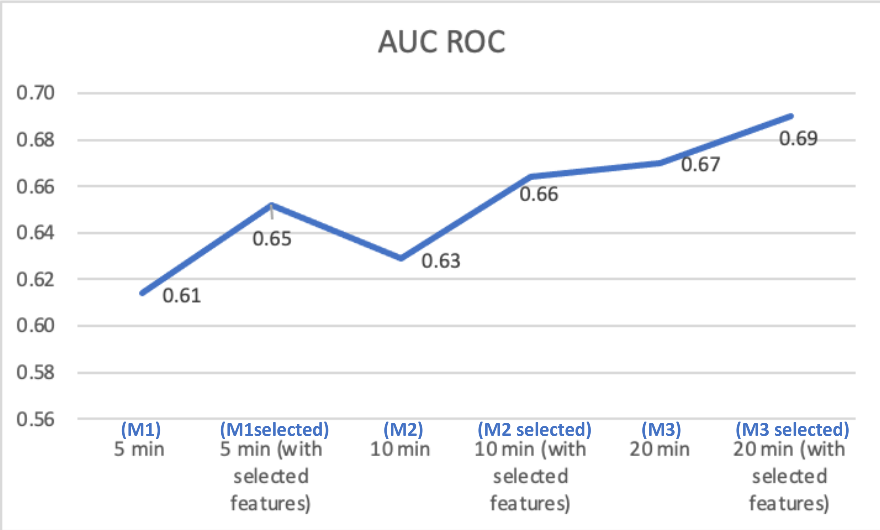



**Figure 2:** AUC score with and without feature selection (k-fold with k=10)

## 4. DISCUSSION AND CONCLUSION

Too early predictions (first five minutes) lack sufficient information, but performance improves with an expanded time window. This aligns with Ulitzsch et al. (2022), who demonstrated enhanced classification performance with larger early-window datasets. Moreover, the study shows limitations. For example, the exploratory nature of this study limits the ability to make theoretical inferences and validate theory-derived indicators (Han et al., 2019). In conclusion, random forest is a robust machine-learning algorithm that is well-suited for analyzing student log files to predict problem-solving success at an early stage of processing. An early understanding of the problem-solving process by using machine learning models and early window log data is helpful to improve problem solving success. However, the type of reasons for less successful problem solving (e.g., low media competences or no interest for the topic at all) should be understood to develop effective interventions.

### References

Brandl, L., Richters, C., Radkowitsch, A., Obersteiner, A., Fischer, M. R., Schmidmaier, R., & Stadler, M. (2021). Simulation-based learning of complex skills: predicting performance with theoretically derived process features. *Psychological Test and Assessment Modeling*, *63*(4), 542-560.

Han, Z., He, Q., & Von Davier, M. (2019). Predictive feature generation and selection using process data from PISA interactive problem-solving items: An application of random forests. *Frontiers in psychology*, *10*, 2461.

He, Q., & Von Davier, M. (2015). Identifying feature sequences from process data in problem-solving items with n-grams. In *Quantitative psychology research* (pp. 173-190). Springer, Cham.

Li, X., Wang, T., & Wang, H. (2017, March). Exploring n-gram features in clickstream data for MOOC learning achievement prediction. In *International Conference on Database Systems for Advanced Applications* (pp. 328-339). Springer, Cham.

Lu, O. H., Huang, A. Y., Huang, J. C., Lin, A. J., Ogata, H., & Yang, S. J. (2018). Applying learning analytics for the early prediction of Students' academic performance in blended learning. *Journal of Educational Technology & Society*, *21*(2), 220-232.

Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & Education*, *143*, 103676. https://doi.org/10.1016/j.compedu.2019.103676.

Ulitzsch, E., Ulitzsch, V., He, Q., & Lüdtke, O. (2022). A machine learning-based procedure for leveraging clickstream data to investigate early predictability of failure on interactive tasks. *Behavior Research Methods*, 1-21.

# A Systematic Review of Self-Regulated Learning Through Multimodal Data and machine learning Integration - Insights from the Grid

**Susanne de Mooij[1], Joni Lämsä[2], Maria Bannert[3], Roger Azevedo[4], Dragan Gasevic[5], Sanna Järvelä[2] & Inge Molenaar[1]**
[1]Radboud University, Netherlands; [2]University of Oulu, Finland; [3]Technical University of Munich, Germany; [4]University of Central Florida, USA; [5]Monash University, Australia
susanne.demooij@ru.nl

**ABSTRACT**: This poster presents a systematic literature review on self-regulated learning (SRL). The aim is to address the challenge of understanding how multimodal data can capture the temporality and sequence of learners' cognitive, metacognitive, affective, and motivational processes (CAMM). In this review, we visualize empirical studies on the self-regulated learning processes, multimodal data, and analysis (SMA) grid—a two-dimensional framework. This grid positions SRL research, revealing relations and potential data stream combinations for CAMM process measurement. We define four analytical approaches in the grid (unimodal, horizontal, vertical, and integrated approaches) to explore interactions between axes and the application of artificial intelligence. The results show a historic shift from an unimodal approach to using multimodal integrated approaches in capturing temporal and sequential characteristics of SRL processes. Identified gaps include limited temporal measures of motivation and affective states, and predominant use of standard statistics in integrated approaches, suggesting the potential for machine learning. This discussion emphasizes the growing role of machine learning in advanced data analytics to comprehend the intricate nature of SRL.

**Keywords**: self-regulated learning, multimodal data, systematic review, artificial intelligence

## INTRODUCTION

In today's fast-paced world, self-regulated learning (SRL) has garnered increased attention, with the hope of empowering students to detect, diagnose, and act upon their own SRL processes more effectively (Molenaar, 2022). SRL, as defined by theory, involves students making conscious choices to achieve learning goals (Winne, 2015). The measurement of SRL has posed challenges in the research field. Multimodal data, advanced data analytics, and the application of AI offer promising avenues to address these challenges. Specifically, interdisciplinary efforts in these fields have utilized SRL theory as a foundational framework, aiming to enhance the measurement and support of SRL processes through the integration of Learning Analytics (LA) and Artificial Intelligence (AI) (Azevedo & Gašević, 2019).

To advance this research line, a comprehensive understanding of the diverse data modalities' ability to non-intrusively capture cognitive, metacognitive, affective, and motivational states of learners over time, collectively known as CAMM processes (Bannert et al., 2017), is crucial. Earlier work by Molenaar et al. (2023), proposed a framework with the two dimensions of *multimodal data* and *CAMM processes* to conceptualise the relations between SRL processes and multimodal data. In this poster, we show the first results of a systematic review investigating how studies map on to this framework showing that different multimodal data are applied to measure CAMM processes and which analytic approaches are used.

## AIMS

This review uses the Self-regulated Learning Processes, Multimodal Data, and Analysis (SMA) grid, delineated along one axis by SRL CAMM processes (Cognition, Affect, Metacognition, and Motivation processes) and on the other axis by multimodal data streams (comprising behavioural, physiological, and contextual data categories). Empirical research in the field of SRL is reviewed and positioned on this grid. The studies are categorized based on their analytical approaches: (1) **Unimodal approach** - one data stream underlying one CAMM process; (2) **Horizontal** - one data stream underlying multiple CAMM processes; (3) **Vertical** - multiple data streams underlying one CAMM process; (4) **Integrated** approach - a combination of data streams and processes.

This grid serves as a visual representation, illustrating the historical evolution of measuring SRL with multimodal data and highlighting how studies have used diverse data analytics techniques in assessing SRL, thereby demonstrating the complexity of the topic at hand.

## METHODOLOGY

We used 3 databases (ERIC, Scopus, Web of Science) to identify peer-reviewed, empirical journal articles and proceedings published in English. Articles ($N$ = 9416) were selected that include search terms such as self-regulated learning or one of the CAMM processes (e.g., motivation) and at least one data stream (e.g., video or chat content). From the 277 studies that were screened fully, 131 studies so far were mapped in the SMA grid according to their SRL process, data stream(s) and analytical approach. These studies were included when they had process-oriented data and analyses.

## FINDINGS

Initial findings are based on the analysis done so far of some of the studies analysed and visualised in Figure 1. The linewidth and size of the circles indicate the frequency of represented studies. The figure reflects perspectives taken in SRL research to detect CAMM processes with the underlying data streams. Most studies have focused on the relation between cognition and metacognition with either a horizontal approach (15% of studies) or an integrated approach (12.5%), but less so on motivational and affective states (< 5% of studies). In addition, Table 1 shows that studies collected more process-oriented data (from 28 in the period 2010-2016 to 94 studies in period 2017-2023) as well as more multimodal data in a vertical and integrated approach.

**Table 1: Comparison of percentages analytical approaches in publication year 2010 and 2022**

| Analytical Approach used by studies | Publication Year 2010-2016 | Publication Year 2017- 2023 |
|---|---|---|
| Unimodal | 28.5% ($N$ = 8) | 22.3% ($N$ = 21) |
| Horizontal | 39.3% ($N$ = 11) | 40.4% ($N$ = 38) |
| Vertical | 21.4% ($N$ = 6) | 8.5% ($N$ = 8) |
| Integrated | 10.7% ($N$ = 3) | 28.7% ($N$ = 27) |


**Figure 1: SMA grid (Self-regulated learning processes, Multimodal data and Analysis) mapping 63 studies**

## CONCLUSION

This systematic review paper introduces a map of the research done in the field of SRL on the self-regulated learning process, multimodal data, and analysis (SMA) grid. This review shows hot spots - i.e., a trend from unimodal to multimodal integrated approaches - as well as gaps in the reviewed studies – i.e., less temporal process measures of student motivation processes. Although more studies are using integrated complex combinations of data streams and modalities where machine learning approaches seem appropriate, the use of standard statistics is still dominant. The remaining challenges where machine learning can help lie in how to align different granularity levels of data streams and combine theory-driven validation criteria (such as think-aloud coding schemes) with data-driven approaches from machine learning.

## REFERENCES

*Molenaar, I. (2022). The concept of hybrid human-AI regulation: Exemplifying how to support young learners' self-regulated learning. Computers and Education: Artificial Intelligence, 3, 100070.*

*Molenaar, I., de Mooij, S., Azevedo, R., Bannert, M., Järvelä, S., & Gašević, D. (2023). Measuring self-regulated learning and the role of AI: Five years of research using multimodal multichannel data. Computers in Human Behavior, 139, 107540.*

*Azevedo, R., & Gašević, D. (2019). Analyzing Multimodal Multichannel Data about Self-Regulated Learning with Advanced Learning Technologies: Issues and Challenges. Computers in Human Behavior, 96, 207–210. https://doi.org/10.1016/j.chb.2019.03.025*

*Winne, P. H. (2015). Self-Regulated Learning. In International Encyclopedia of the Social & Behavioral Sciences: Second Edition. https://doi.org/10.1016/B978-0-08-097086-8.25091-5*

# A Data-Centric Personalized Learning Technology Solution to Accelerate Early Math Skills of Young Learners

**Linlin Li, Mingyu Feng, Natalie Brezack**
WestEd
lli@wested.org, mfeng@wested.org, nbrezac@wested.org

**Hee Jin Bang**
Age of Learning Inc.
heejin.bang@aofl.com

**ABSTRACT**: Recognizing the importance of building an early foundation to support lifelong learning and success, schools and districts seek tools and approaches that engage active learners and enhance personalized learning. Despite the availability of proven early mathematics instructional programs, teachers struggle to implement individualized learning plans for each student. Technology-based programs can reduce this burden on teachers and support students' personalized learning. In this poster, we present a highly engaging educational technology innovation targeting early elementary students – *My Math Academy* – that leverages data to drive individual learning paths and uses learning analytics to inform classroom instructions.

**Keywords**: Personalized Learning Technology, Early Math, Game-based Program, Data Analytics

## 1 INTRODUCTION

Despite the availability of proven early mathematics instructional programs, teachers struggle to implement individualized learning plans for each student. Technology-based programs can reduce this burden on teachers and support students' personalized learning. Game-based programs offer additional benefits along with increasing consistent delivery of technology-based and individualized learning to students; games foster engagement and motivation by providing interactivity, adaptive challenges, and ongoing feedback. They offer safe environments for failure, encouraging learners to take risks, explore, and try new things (Hoffman & Nadelson, 2010). Furthermore, digital games can track, assess, and generate data on student learning, providing information that helps teachers better understand individual students' learning needs and plan instruction accordingly (Goddard et al., 2015).

## 2 MY MATH ACADEMY PROGRAM

Highly engaging educational technology innovations targeting early elementary students, such as *My Math Academy*, can close the achievement gap and prepare students for success in mathematics. *My Math Academy* uses a **mastery-based personalized learning** approach that consists of game-based activities with adaptive learning trajectories, performance dashboards that help teachers support students' learning, and offline activities that extend in-game learning experiences. It respects learner variability by differentiating instruction and providing appropriate feedback, ensuring that children master each topic before moving on. Students receive pre-assessments, instruction, feedback, and corrections. *My Math Academy* also integrates **evidence-centered design** and enables the estimation of students' competency levels via in-game learning data (Mislevy, 2011).

The teacher dashboard provides teachers with real-time student progress data and supplemental materials to support instructional decision-making (Figure 1). The dashboard provides an overview of the class, which can be filtered into teacher-created groups. Teachers can view individual student's progress on each learning objective and access recommendations based on their in-game performance (i.e., ready to learn; needs for review, reinforcement, or intervention). The dashboard also suggests activities for students stuck on learning objectives, providing teachers with data about students' current proficiency and learning trajectories to help them better tailor instruction.


**Figure 1: Teacher Dashboard Displays**

## 3 STUDIES AND FINDINGS

Two pilot randomized control trial (RCT) studies (n = 958; n = 428) conducted across several districts in California, USA, showed that transitional-kindergarten, kindergarten, and first grade students who used *My Math Academy* had statistically significantly higher scores on the post-test (Bang et al., 2022; Thai et al., 2021). Additionally, a quasi-experimental study (n = 976) conducted during a school year disrupted by the pandemic (SY20-21) with pre-kindergarten students in a Title I school district in Texas, USA showed that 98% of those who used *My Math Academy* regularly ended the school on track in math on the state-administered assessment (Bang & Thai, 2022). Another quasi-experimental study (n = 3,445) conducted across high-needs voluntary pre-kindergarten centers in Florida, USA in spring 2023 showed that using *My Math Academy* helped close the gap between students who used the program and those who did not by 84% on the state-administered assessment (Bang & Setoguchi, 2023). Across these studies, educators reported that *My Math Academy* had a positive impact on students' enjoyment, interest, and self-confidence in learning math. They also recognized the value of the program for personalizing learning and advocated for its continued use.

Currently, a large-scale RCT involving more than 600 kindergarten students and 35 teachers from 15 schools in 5 public school districts in the USA is underway to examine the efficacy of *My Math Academy*. In the pilot study conducted in spring 2023, interviewed teachers reported valuing *My Math Academy*'s engaging, differentiated, and adaptive features to support individualized learning:

> What I mostly liked about it was that it, you know, did the assessments and kind of differentiated for me because as a teacher with 25 kids… it's a lot to differentiate and meet each child where they are. So it was nice that My Math Academy kind of did that part for you.
>
> It's nice to have My Math Academy to meet kids where they are, especially those kids who don't meet the standards for a particular skill… there were children who were doing, like I said, just ordering numbers, which we covered towards the beginning of kindergarten... it was nice to see that, although I had to keep moving on with the curriculum, My Math Academy was still bringing them back to that foundation of putting numbers in order, things like that.

Teachers in the pilot study found the learning analytics from the teacher dashboard to be informative, surprising, and useful for guiding classroom instruction:

> It gives me the valuable tools that I need as an educator to be able to see what their strengths, what their weaknesses are, and where they need more support.
>
> I was surprised to see how advanced like my one student that was on those second grade levels... my low ones, there were some things that I thought they had mastered... [but] I knew [from the Teacher Dashboard] they needed more practice.

## 4 CONCLUSIONS AND FUTURE WORK

*My Math Academy* provides an example of using effective design strategies to empower practical, theoretically-sound technology-based programs featuring playful engagement, learning in context, and formative assessments. The results from multiple studies indicated that *My Math Academy* was positively and significantly associated with gains in students' math skill as well as their interest and self-confidence in learning math. Developers of technology-based programs can further leverage gameplay in their designs to create interactive games that encourage meaningful, active, child-centered learning that engages users and allows for continual growth. Researchers are examining how teachers' effective use of dashboards to provide personalized instruction results in higher student engagement and performance in math.

## REFERENCES

Goddard, Y., Goddard, R., & Kim, M. (2015). School instructional climate and student achievement: An examination of group norms for differentiated instruction. *American Journal of Education*, *122*(1), 111–131.

Hoffman, B. & Nadelson, L. (2010). Motivational engagement and video gaming: A mixed methods study. *Educational Technology Research and Development, 58(*3), 245-270.

Mislevy, R. J. (2011). *Evidence-centered design for simulation-based assessment (CRESST Report 80*0). University of California, Los Angeles, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Thai, K., Bang, H.J., & Li, L. (2021). Accelerating early math learning with research-based personalized learning games: A cluster randomized controlled trial. *Journal of Research on Educational Effectiveness.* https://doi.org/10.1080/19345747.2021.1969710.

# Managerial and Strategic Learning Analytics

**Ronie Navon**
The Open University of Israel
ronie@technion.ac.il

**Yoram M Kalman**
The Open University of Israel
yoramka@openu.ac.il

**ABSTRACT**: This poster addresses an interesting gap between the *current* and *potential* applications of learning analytics (LA) at educational institutions. A review of 133 full and short conference papers presented at LAK in 2022 and 2023 reveals that 95% of the papers dealt with aspects of LA limited to the "classroom-level", addressing the needs of the students and their instructors, or dealt with issues which could neither be classified as "classroom-level", nor as "institutional". In contrast, only 5% of the papers addressed issues related to understanding and improving teaching and learning from a managerial and strategic organizational perspective.

**Keywords**: Learning Analytics, Educational Institutions, Stakeholders

## 1 INTRODUCTION

Learning analytics (LA) is "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (What is Learning Analytics?, n.d.). The goal of this poster is to demonstrate a thought-provoking imbalance in LA research. Specifically, we demonstrate that although the quality of, and impact on, teaching and learning (T&L) are influenced by many activities and decisions, a significant majority of LA research and practice focus directly on informing instructors about their students (Dawson et al., 2023) and much less attention is paid to the activities and decisions of other stakeholders in educational institutions (EIs) (Herodotou et al., 2019; Lin et al., 2023).

We use the term "classroom-level" to denote activities and decisions related to the student, the instructor, and the interactions between them that take place in class and outside of it. Accordingly, we use the term "institutional" to denote managerial and strategic activities taken by other stakeholders in the institution, and even beyond the individual institution. This poster demonstrates the imbalance between the extensive amount "classroom-level" LA research, vs. the relatively limited amount of "institutional" LA research that deals with activities and decisions taken outside the classroom but impact T&L no less. Some examples of such activities and decisions include resource allocation to classes (e.g., class size), curriculum design, admissions criteria, strategic goal setting, etc. (Dawson et al., 2023; Herodotou et al., 2019). In the case of higher education, these are decisions mainly taken by department heads, provosts and presidents, administrative staff, etc. Furthermore, the focus on the "classroom-level" often leads to overlooking many sources of data such as students enrollment over time, student activity in the libraries, distribution of grades across courses and departments, students engagement with learning resources, student satisfaction surveys, etc. (Dawson et al., 2023; Herodotou et al., 2019).

To substantiate the claim that LA research primarily focuses on the classroom-level, we reviewed 133 short and long papers published in the proceedings of LAK 22 and LAK 23.

## 2 METHOD

An exploratory review of the 133 LAK papers was carried out in several steps. The initial step was a preparatory review of a sample of 27 papers by one of the researchers. Consequently, the researcher suggested three classifications relating to the context in which the LA were researched: "classroom-level", "institutional" and "other". As described above, the term "classroom-level" denoted activities and decisions related to the student, the instructor, and the interactions between them, while the term "institutional" denoted all other contexts which relate to the EI and even beyond (e.g. regulating bodies and government agencies). The "other" classification was assigned to papers that did not discuss applications of LA relating to the EI, such as LA related to software development, machine learning-based LA, computer vision for behavioral research, LA related research methods, students' self-efficacy, collaboration patterns and quality, etc. All of these did not research nor discuss issues related to decisions or activities either in the classroom or by other decision makers at EIs. Importantly, the classification of papers as "institutional" was given even to papers that only *discussed* implications beyond the classroom. This classification was tested by the second researcher who randomly sampled 10 of the 27 papers. The classifications were almost identical, with one disagreement, classifying one paper that was classified as "other" by one researcher, and as "classroom-level" by the other. Since the focus of the poster is identifying "institutional" papers, the first researcher continued with the classification of the remaining 114 papers. Since the frequency of the "institutional" papers was low (less than 10%), the second researcher verified these classifications by randomly sampling 20 papers that were classified "classroom-level" or "other", as well as *all* of papers that were classified "institutional". The agreement at that stage was, again, almost full, with one minor disagreement about the classification of one paper as "other" or "classroom-level", which was discussed between the researchers and agreed upon.

## 3 RESULTS

In LAK 22, 4 (6%) of the papers were classified "Institutional", while 44 (71%) were classified "classroom-level" and 14 (23%) as "other". In LAK 23, 2 (3%) papers were classified "Institutional", 62 (87%) as "classroom-level" and 7 (10%) as "other". The results are depicted in Figure 1.

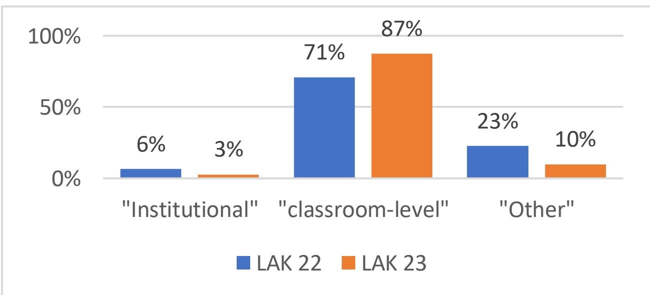



**Figure 1: LAK 22-23 papers by classification**

# 4 DISCUSSION

This poster presents evidence for a gap between the current and the potential applications of LA to decisions and activities that take place *outside* of the classroom. Given that such managerial and strategic level decisions can directly impact T&L, we call on the research community to consider this additional perspective in their research. Anecdotally, while performing the classification, we came across a number of papers whose findings *could* have interesting and useful implications beyond the classroom, but these were not mentioned.

Why does this gap exist? One possible explanation is that most LA researchers are faculty members extensively involved in teaching and are thus most interested in "classroom-level" research and implications. Another explanation could be that some view analytics related to managerial and strategic decisions as *managerial* analytics rather than *learning* analytics. But, given that the purpose of LA is "understanding and optimizing learning and the *environments in which it occurs*" (see Section 1, above. Emphasis added), and since managerial and strategic decisions taken outside the classroom, extensively impact the environments in which learning takes place, as well as the learning itself – more attention to the "institutional" category could have a significant positive impact on the quality of T&L.

Finally, we will briefly mention that a stronger focus on measuring the impact of decisions and actions taken outside the classroom on T&L could improve overall performance of educational institutions, but that this also raises novel ethical issues (Tzimas & Demetriadis, 2021). Many of these issues have to do with the need to train additional stakeholders about potential lapses and negative consequences that could unintentionally arise from uninformed usage of data and analytics.

# REFERENCES

Dawson, S., Pardo, A., Salehian Kia, F., & Panadero, E. (2023). An Integrated Model of Feedback and Assessment: From fine grained to holistic programmatic review. *13th International Learning Analytics and Knowledge Conference*, 579–584.

Herodotou, C., Hlosta, M., Boroowa, A., Rienties, B., Zdrahal, Z., & Mangafa, C. (2019). Empowering online teachers through predictive learning analytics. *British Journal of Educational Technology*, *50*(6), 3064–3079.

Lin, J., Dai, W., Lim, L.-A., Tsai, Y.-S., Mello, R. F., Khosravi, H., Gasevic, D., & Chen, G. (2023). Learner-centred analytics of feedback content in higher education. *13th International Learning Analytics and Knowledge Conference*, 100–110.

Tzimas, D., & Demetriadis, S. (2021). Ethical issues in learning analytics: A review of the field. *Educational Technology Research and Development*, *69*, 1101–1133. https://doi.org/10.1007/s11423-021-09977-4

What is Learning Analytics? (n.d.). What is Learning Analytics? *Society for Learning Analytics Research (SoLAR)*. Retrieved November 9, 2023, from https://www.solaresearch.org/about/what-is-learning-analytics/

# Bridging learning science and learning analytics: Self-Regulation Learning support (SRL-S) rubric

**Slaviša Radović**
Center of Advanced Technology for Assisted Learning and Predictive Analytics (CATALPA), FernUniversität in Hagen
slavisa.radovic@fernuni-hagen.de

**Niels Seidel**
Center of Advanced Technology for Assisted Learning and Predictive Analytics (CATALPA), FernUniversität in Hagen
niels.seidel@fernuni-hagen.de

**ABSTRACT**: The integration of advanced learning analytics to support students' self-regulation into higher education has brought the huge mismatch between the principles of learning science (such as the SRL model), the design of learning analytics technologies, and their evaluation in empirical research. This poster presentation aims to address this gap by introducing the SRL-S rubric, firmly grounded in Zimmerman's theoretical model and learning analytics features that have demonstrated substantial efficacy in empirical studies. Aligned with the three phases of Self-Regulated Learning, the rubric establishes connections between assessment criteria for self-regulated learning support and performance levels within learning environments (Limited, Moderate, and Advanced). By employing the rubric, educators and researchers can gain insights into the extent of implemented SRL approaches and further improve missing SRL support. This poster presents the rubric's framework, its developmental process, validation analyses, and concludes by discussing its pivotal role in advancing self-regulated learning amidst the current educational landscape.

**Keywords**: Self-regulated learning, rubric, learning analytics, learning science.

## 1 SUPPORTING SRL IN HIGER EDUCATION

Over the past two decades, technology enhanced learning environments in higher education has gained significant prominence and transformed education process (Jin et., 2023). These learning environments are often characterized by higher freedom of students to choose when, what and where they want to learn, reduced teachers' presence, and greater students' autonomy over their own learning (Breitwieser et al., 2023; Radović & Seidel, 2023). In this educational context, where learners often navigate the learning process with greater ownership, self-regulated learning (SRL) skills have emerged as essential prerequisites for effective, efficient, and enjoyable learning experiences (Jin et., 2023; Zimmerman, 2000). Consequently, various advanced learning technologies and tool have been developed on the ground of learning analytics, and furthermore adapted to support students developing these skills. These include the utilization of learning analytics dashboards, support for goal setting, incorporation of self-assessment features, guidance for student reflection, and provision of personalized recommendations (Edisherashvili et al., 2022; Jivet et al., 2017).

While some of these technologies and tools have proven beneficial for the learning process, the results from empirical studies are not always conclusive and positive (Radović, 2023). As the technological landscape continues to evolve in support of SRL (Edisherashvili et al., 2022), it becomes increasingly evident that the nature of this support is complex and cannot be simplified into a binary or dichotomous concept. Instead, SRL support encompasses a wide spectrum of advanced learning technologies that extend beyond a simplistic categorization of whether a learning environment supports or does not support SRL (Radović et al., 2024). Rather, it is a range, from limited (or even no support) to advanced support.

Literature reviews also acknowledge a huge mismatch between learning science (SRL model), learning analytics (design of technologies), and evaluation in empirical articles (Pérez-Álvarez et al., 2018). For example, many past initiatives focus on partial pedagogic support (e.g. only on learning dashboard, or implementing only self-assessment tasks), rather than utilizing a comprehensive support for all phases of SRL (Jivet et al., 2017; Radović et al., 2024). Placing emphasis on certain aspects of the SRL process while ignoring others is insufficient and even hindering the learning process (for instance, praising the monitoring phase while neglecting the reflection phase) (Radović et al., 2024).

## 2 SELF-REGULATION LEARNING SUPPORT (SRL-S) RUBRIC

The two mentioned challenges impede the understanding of provided SRL support in specific cases and the ability to compare different developments on a standardized scale (Radović & Seidel, 2024). Since there is no standardized scale, in this poster presentation we present our newly developed SRL-S rubric, designed to assess the degree of SRL support available within technology enhanced learning environments (Radović & Seidel, 2024). We combined theoretical literature and proven empirical results to create a stronger bond between learning science and learning analytics, to empower researchers in their endeavors of SRL support development.

| Phase | Process | Limited SRL support (1) | Moderate SRL support (2) | Advanced SRL support (3) |
|---|---|---|---|---|
| Forethought | *F1. Goal Setting* | Students acquire course goals predefined by the teacher, they do not have the option to set or modify their goals within learning environment, nor can they easily access goal related performance indicators. | While students still lack the capability to set or change learning goals themselves in the learning environment itself, however they receive detailed insights about their learning concerning the course's goal. | Students enjoy the flexibility to choose from a range of learning goals (which may include course mastery or just passing) or to set custom goals (content or performance related). Additionally, students are provided with details related to the chosen goal. |
| Self-Reflection | *S2. Causal Attribution* | Students are offered a limited resources to reflect (e.g., knowledge tests and related rubrics). They are not guided nor supported how to reflect on performance or how to evaluate factors of failure. | Students are asked to think about their performance when self-assessing tasks' solutions against criteria. This level of support encourages students to consider the factors that influenced their failures. | Learning environment includes prompted critical reflection tasks after major learning events or learning units. These tasks ask students to think– about their performance, their strengths and weaknesses, as well as to assess their progress toward their goals. |

**Figure 1.** ***The glimpse of the SRL-S rubric shows only two SRL criteria (F1 from Forethought and S2 from Self-Reflection phase ) with corresponding performance levels.***

The rubric criteria are structured according to the phases and subprocesses of SRL. This framework is firmly rooted in educational theory, particularly drawing from Zimmerman's theoretical model and other seminal articles (Zimmerman, 2000). The performance levels (limited, moderate, and advanced) have been derived from empirical research and systematic literature studies, such as those conducted by Jivet et al. (2017) and Pérez-Álvarez et al. (2018). Advanced technological tools were categorized based on the phases and processes of SRL they primarily enhance. Subsequently, we organized these tools into three distinct levels for each SRL subprocess, thereby establishing clear and specific standards for each criterion. For an in-depth description of the development process and validation analysis (teacher interrater and intrarater reliability), please refer to Radović and Seidel (2024) and demo https://catalparesearch.github.io/srl-support-rubric/.

# 3 DISCUSSIONS

In pursuit of addressing recent challenges in SRL field, the rubrics provide comprehensive assessment criteria and offer detailed definitions of performance levels that span from Limited to Advanced SRL support. While we amalgamate theoretical and empirical literature considering learning analytics, we acknowledge that other various factors beyond learning technologies, such as pedagogical design, teachers' expertise and experience, and delivery modes (e.g., fully online, hybrid, etc.), could also support students' regulation. However, that was beyond our current exploration.

By employing the rubric, educators and researchers can gain objective assessment of the extent of implemented SRL support including learning analytics instruments in their courses and learning environments. Moreover, based on results (missed technologies and features), they can further develop students' SRL support and better support students on their journey towards becoming self-regulated learners. Finally, this rubric can become a basis for measuring SRL activities (usage of SRL S).

# REFERENCES


Breitwieser, J., Nobbe, L., Biedermann, D., & Brod, G. (2023). Boosting self-regulated learning with mobile interventions: Planning and prompting help children maintain a regular study routine. Computers & Education, 205, 104879. https://doi.org/10.1016/j.compedu.2023.104879

Edisherashvili, N., Saks, K., Pedaste, M., & Leijen, L. (2022). Supporting Self-Regulated Learning in Distance Learning Contexts at Higher Education Level: Systematic Literature Review. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.792422

Jin, S. H., Im, K., Yoo, M., Roll, I., & Seo, K. (2023). Supporting students' self-regulated learning in online learning using artificial intelligence applications. International Journal of Educational Technology in Higher Education, 20(1). https://doi.org/10.1186/s41239-023-00406-5

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice. In European conference on technology enhanced learning (pp. 82-96). Springer, Cham.

Pérez-Álvarez, R., Maldonado-Mahauad, J., & Pérez-Sanagustín, M. (2018). Tools to support self-regulated learning in online environments: literature review. In European Conference on Technology Enhanced Learning (pp. 16–30). Springer, Cham.

Radović, S. (2023). Is it only about technology? The interplay between educational technology, teaching practice, and students' learning activities. Journal of Computers in Education, DOI:10.1007/s40692-023-00277-9.

Radović, S., Seidel, N., Menze, D., & Kasakowskij, R. (2024). The nuances of self-regulated learning: Investigating the impacts of different levels of students' regulation support in distance higher education. *(Accepted for publication in Computers & Education)*

Radović, S., & Seidel, N. (2024). Beyond the dichotomy: the SRL-S rubric for assessing the extent of educational technology's support of self-regulated learning in higher education. *(Submitted for publication)*

Zimmerman, B. J. (2000). Attaining self-regulation: a social cognitive perspective. In Handbook of Self-Regulation, eds M. Boekaerts, P. R. Pintrich, and M. Zeidner (CA: Academic Press), 13–40.

# A Sequential and Clustering Analysis of Preservice Teachers' Self-Regulation in Learning Complex Professional Knowledge

**Lingyun Huang**
Department of Curriculum and Instruction, The Education University of Hong Kong
lingyunhuang@eduhk.hk

**ABSTRACT**: This study explores the relations between preservice teachers' self-regulation and their development of technological pedagogical content knowledge (TPACK). Twenty-eight participants were invited to develop TPACK by designing a technology-infused lesson in a computer-based learning environment. Their self-regulation processes were recorded, coded, and analyzed through think-aloud data and sequential clustering analysis. The results show two distinct two regulation process patterns. One group, labeled the low-regulation group, had a shorter sequence length and dominantly enacted elaboration activities. In contrast, the other had longer sequences, engaged in diverse self-regulatory activities, and was labeled the high-regulation group. Relating to TPACK development evaluated by the quality of lesson plans, the results indicate that the participants in the high-regulation group outperformed their counterparts in the low-regulation group. The findings echo the previous evidence and provide implications for practitioners about the importance of preservice teachers' self-regulation in TPACK development.

**Keywords**: TPACK, Sequential and clustering analysis, think-aloud protocols

## 1 INTRODUCTION

TPACK emphasizes that technology use should be aligned with pedagogy, content, learners, and learning context to address issues in how particular topics are difficult to be understood by learners or difficult to be represented by teachers (Angeli & Valanides, 2009). While developing TPACK, preservice teachers should be aware of the ill-defined nature of TPACK and adapt their technology use according to the learning context and the nature of the problem (Huang et al., 2023). Previous research reinforces the importance of self-regulated learning (SRL) for TPACK development (e.g., Huang et al., 2023). Preservice teachers are supposed to learn how to monitor and regulate their efforts to help them progress toward TPACK development (Kramarski & Michalsky, 2010). nBrowser (Poitras et al., 2017) is a technology-rich learning environment designed for preservice teachers to scaffold their SRL and facilitate their learning of TPACK in addressing complex scenario-based authentic instructional cases. The SRL process consists of numerous specific regulatory activities that can be indicated by actual events (Greene & Azevedo, 2010). nBrowser logs preservice teachers' actions within the TPACK context to identify preservice teachers' SRL processes. This study proposes the sequential clustering method to mine self-regulatory process patterns and aims to explore whether a positive correlation can be obtained between SRL and TPACK, echoed in the existing literature (e.g., Huang et al., 2023). The two research questions are (1) whether the sequential clustering method can generate two clusters indicative of preservice teachers' SRL patterns in the TPACK context and (2) whether there is a difference in TPACK between the two clusters. We assume two clusters can be identified: a high and a low level of SRL. For the second question, we hypothesize that there are statistically significant differences in TPACK between the two clusters. More specifically, the high SRL cluster would have better TPACK abilities than the low one (Huang et al., 2023).

## 2 METHODS

Twenty-eight participants (female = 24) involved in this study were preservice teachers from a normal university in China. The participants' mean age was 20.86 years (SD = .82). They were asked to design an English lesson and use appropriate technology to facilitate their teaching in nBrowser. At first, the participants received a consent form, an introduction video of the nBrowser, and instructions on thinking aloud. The experimenter met the participants individually, introduced the study, and clarified the participants' concerns. Then, the participants had 45 minutes to complete the task. Then, the participants had 45 minutes to complete the task, that is, to design a technology-infused English lesson on the topic of the Canadian Tulip Festival. While doing the task, the participants were asked to verbalize their thoughts, which were audio-recorded. The system automatically logged participants' actions and lesson plans.

## 3. ANALYSIS

Participants' audio recordings were transcribed into think-aloud protocols and coded using a scheme (Huang et al., 2023) adapted from the micro-SRL coding scheme (Greene & Azevedo, 2010). Two researchers followed the coding process to complete the task: trail-discussion-individual coding-peer review-adjustment. The first author of this paper audited the process and resolved disagreements during and after coding. Of the total 1156 codable segments, there were 124 disagreements. Hence, there is no need for inter-rater reliability as every codable segment was evaluated by two separate researchers, with any differences addressed through discussions (Greene & Azevedo, 2010). Lesson plans that reflected on participants' TPACK were analyzed based on the rubric (Huang & Lajoie, 2021)

The study utilizes TraMineR, which mines the sequences of states or events using hierarchical clustering algorithms and measures the similarities and distances between multiple sequences (Gabadinho et al., 2011). In this study, we used the individual SRL events to mine the sequences of preservice teachers' SRL in TPACK development. Using visualization functions, we can create the dendrogram graph visualizing the distinct clusters with the involved members.

## 4. RESULTS, DISCUSSION, AND CONCLUSION

For the first question, the dendrogram result suggests two clusters as the best result, with an agglomerative coefficient of 0.78. Cluster 1 has 17 participants (60.71%) and contains shorter sequences ranging from 16 to 35, whereas Cluster 2 (n = 11 39.29%) contains longer sequences, ranging from 48 to 80. Given that, we label Cluster 1 as a Low SRL group and Cluster 2 as a High SRL group. Moreover, the result shows that the participants in the Low SRL group dominantly enacted the Elaboration event. In comparison, the participants in the High SRL group engaged in diverse events such as Execution, Monitoring, and Elaboration. The second question is to compare the TPACK mean between two SRL clusters. The result of the independent t-test shows that TPACK outcomes ($M = 3.91$, $SD = 1.70$) in the high SRL group were statistically significantly higher than that of the low SRL group ($M = 2.53$, $SD = 1.66$), evidenced by $t(26) = 1.38$, $p < .05$.

In the study, we investigated how preservice teachers regulated their learning of TPACK in a computer-based platform and how their SRL sequences affected learning outcomes. The results of the sequence-based clustering identified two SRL clusters that contained significant differences in SRL sequential patterns. One of the distinctions is that the Low SRL group frequently enacted Elaboration events, suggesting that preservice teachers in this group spent more time explaining their behaviors and decisions. For the High SRL group, the sequences exhibit that members were more regulated. They monitored resource searching and saving activities, justifying what and why to save and evaluating whether the resources were useful for goals. Such sequences are aligned with SRL models and are effective for better outcomes (Lim et al., 2021). The findings show that preservice teachers' SRL can positively correlate with TPACK learning. Those who exhibited higher SRL had better achievements. The findings confirm the positive relationship between SRL and TPACK (Huang et al., 2023).

Nevertheless, we understand that the participants from the same major and university may affect the generalizability of our research findings. Future research can include preservice teachers from different majors, such as STEM subjects, and various universities. Despite these limitations, this study has several implications. The current study highlights the sequence-based clustering approach, providing a deeper insight into how individual SRL events are connected and arranged. This further demonstrates that some SRL sequences are beneficial for learning, and these patterns are consistent in different contexts and domain learning (Lim et al., 2021). Thus, future research can consider the design of prompts to support SRL sequences rather than individual activities in the TPACK context.

## REFERENCES

Angeli, C., & Valanides, N. (2009). Epistemological and methodological issues for the conceptualization, development, and assessment of ICT–TPCK: Advances in technological pedagogical content knowledge (TPCK). *Computers & Education*, *52*(1), 154–168. https://doi.org/10.1016/j.compedu.2008.07.006

Gabadinho, A., Ritschard, G., Studer, M., & Muller, N. S. (2011). *Mining sequence data in R with the TraMineR package: A user's guide*. Laboratory of Demography, University of Geneva. http://mephisto.unige.ch/pub/TraMineR/doc/TraMineR-Users-Guide.pdf

Greene, J., & Azevedo, R. (2010). The measurement of learners' self-regulated cognitive and metacognitive processes while using computer-based learning environments. *Educational Psychologist*, *45*(4). https://doi.org/10.1080/00461520.2010.515935

Huang, L., Doleck, T., Chen, B., Huang, X., Tan, C., Lajoie, S. P., & Wang, M. (2023). Multimodal learning analytics for assessing teachers' self-regulated learning in planning technology-integrated lessons in a computer-based environment. *Education and Information Technologies*. https://doi.org/10.1007/s10639-023-11804-7

Huang, L., & Lajoie, S. P. (2021). Process analysis of teachers' self-regulated learning patterns in technological pedagogical content knowledge development. *Computers & Education*, *166*, 104–169. https://doi.org/10.1016/j.compedu.2021.104169

Kramarski, B., & Michalsky, T. (2010). Preparing preservice teachers for self-regulated learning in the context of technological pedagogical content knowledge. *Learning and Instruction*, *20*(5), 434–447. https://doi.org/10.1016/j.learninstruc.2009.05.003

Lim, L., Bannert, M., van der Graaf, J., Molenaar, I., Fan, Y., Kilgour, J., Moore, J., & Gašević, D. (2021). Temporal assessment of self-regulated learning by mining students' think-aloud protocols. *Frontiers in Psychology*, *12*, 749749. https://doi.org/10.3389/fpsyg.2021.749749

Poitras, E., Doleck, T., Huang, L., Li, S., & Lajoie, S. P. (2017). Advancing teacher technology education using open-ended learning environments as research and training platforms. *Australasian Journal of Educational Technology*, *33*(3). https://doi.org/10.14742/ajet.3498

# Designing a Teacher Dashboard to Support Primary School Teachers' Direct Strategy Instruction and Students' Self-regulated Learning

**Melis Dülger*, Anouschka van Leeuwen, Jeroen Janssen, Liesbeth Kester**
Utrecht University, The Netherlands
m.dulger@uu.nl*

**ABSTRACT**: Self-regulated learning (SRL) is crucial for students' lifelong learning skills, encompassing planning, monitoring, and control. In the Netherlands, adaptive learning technologies (ALTs) are widely used in math education in primary schools, assisting students with feedback and difficulty adjustments based on students' levels. However, students still need to invest effort and monitor their progress. Teachers play a vital role in teaching SRL strategies. Direct strategy instruction is an effective way to teach these strategies. Yet, existing SRL dashboards lack focus on strategy instruction and theory-driven design. To address these gaps, we aimed to create a classroom-level teacher dashboard through an iterative co-design process based on teacher input and theoretical foundations. To achieve this, we conducted two rounds of interviews with the focus of investigating relevant SRL indicators, teachers' design preferences, and evaluation and optimization of low-fidelity prototypes based on these findings. This study sets the groundwork for future SRL dashboards that enhance teachers' strategy instruction.

**Keywords**: self-regulated learning, direct strategy instruction, teacher dashboard, K-12 education, adaptive learning technologies.

## 1 INTRODUCTION

Self-regulated learning (SRL) is learners' ability to plan, monitor, and control their learning. The COPES model of SRL (Winne & Hadwin, 1998) outlines four phases: task definition, goal setting and planning, enactment, and adaptations. In the Netherlands, adaptive learning technologies (ALTs) developed for math education provide feedback and adjust problem difficulty, providing external regulation. However, students still need strong SRL skills to apply effort and maintain accuracy (Molenaar et al., 2021). For younger students, teachers' instruction of SRL strategies plays an important role in developing SRL skills. Teachers can instruct these strategies through direct instruction, which was found to improve students' math learning (Kistner et al., 2010). Monitoring and identifying needs of students timely and accurately might be challenging for teachers due to crowded classrooms and different learning paths of students. Classroom-level teacher dashboards analyze and report aggregated information regarding students' learning processes in the class, thereby facilitating teachers' monitoring (Van Leeuwen et al., 2015). However, most of these dashboards were not developed to provide information on students' SRL (Wiedbusch et al., 2021) and none of these dashboards targeted teachers' direct instruction of SRL strategies. In addition, most existing teacher dashboards lack the theoretical grounding (Verbert et al., 2020). To address these gaps, we aim to develop a theory-based classroom-level teacher dashboard to support students' SRL using an iterative co-design approach involving primary school teachers. The designed dashboard will be tested and improved in later studies. This study describes initial design steps.

# 2 METHODOLOGY AND PRELIMINARY RESULTS

## 2.1 First Round of Interviews and the Development of Low-fidelity Prototypes

We first conducted interviews to uncover relevant indicators of SRL and to understand primary school teachers' preferences for a classroom-level dashboard that provides information about their students' SRL. Ten upper-level Dutch primary school teachers with experience with ALTs participated in interviews lasting between 30 to 45 minutes. To engage the teachers effectively, we utilized storyboards (Hanington & Martin, 2012). While four of the storyboards illustrated four SRL phases, three storyboards illustrated situations requiring teachers' instructional decisions, as we wanted to investigate relevant SRL indicators and teachers' ideas on direct instruction of SRL strategies. After the storyboard presentation, we posed three reflective questions to gain deeper insights into teachers' instructional strategies and design preferences. Based on our analysis of the interviews, we found that for the task definition phase of SRL, teachers highlighted the importance of monitoring classroom information on students' prior knowledge, learning gaps, and motivation. Most teachers talked about students' prior knowledge aligned with their learning gaps. Concerning the goal-setting and planning phase, teachers expressed the value of gaining insights into students' set and achieved goals to get an overview of their SRL. A few teachers suggested seeing the percentage of learning goals achieved by students while working with ALTs would be beneficial. For the enactment phase, teachers discussed the significance of monitoring students' failed and achieved goals, tracking the number of correct and incorrect answers, and observing overall student growth. Regarding the adaptation phase, teachers expressed interest in understanding where students encounter difficulties and how they navigate challenges to gain an overview of their working strategies. Teachers also acknowledged the use of dashboard information for various purposes, including conversations with the class, lesson preparation, improvement and reshaping of teaching strategies, creating subgroups for instruction, and planning. Throughout the interviews, teachers provided suggestions for the teacher dashboard design and representation of textual and visual elements. Based on the insights obtained from this initial phase and in alignment with SRL theory, we developed two low-fidelity prototypes demonstrating aggregated information on students' SRL. Taken together, we have created the task definition (1), goal-setting (2), enactment (3), and adaptation (4) widgets as shown in Figure 1 to visualize students' learning processes. One of the prototypes included information at the classroom, group, and individual levels, while the other one primarily focused on classroom and group level information.

## 2.2 Second Round of Interviews

Subsequently, we conducted a second round of interviews with eleven upper-level Dutch primary school teachers. The objective of these interviews was to assess the clarity and usability of the dashboard and the information provided in the form of widgets, gather teachers' preferences regarding aggregation of the information, and refine the design. These interviews were complemented with the presentation of four classroom scenarios, each linked to specific widgets on the dashboard, and related questions designed to evaluate the clarity and usability of the information. We also asked additional questions to gather teachers' preferences and suggestions on the dashboard information and design. Our preliminary findings indicated that teachers generally found task definition and enactment widgets clear, but some struggled with interpreting the flag in the goal-setting widget. The adaptations widget was not very clear due to unfamiliar terminology. While teachers saw the information as actionable for instruction, they mostly did not specifically mention strategy instruction. Teachers favored the general overview on the right but desired more detailed individual information on the left prototype, possibly in a separate tab. They also expressed a need for suggestions on how to use the information in the class. These results will serve as the basis for our medium-fidelity prototype, which will undergo testing in a lab study using vignettes simulating actual learning scenarios. At the time of LAK24, study results will be presented along with the prototypes.


**Figure 1: Classroom, group, and individual level dashboard prototype (left), classroom and group level dashboard prototype (right)**

## REFERENCES


Hanington, B., & Martin, B. (2012). *Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions.* Quarto Publishing Group USA. http://ebookcentral.proquest.com/lib/uunl/detail.action?docID=3399583

Kistner, S., Rakoczy, K., Otto, B., Dignath-van Ewijk, C., Büttner, G., & Klieme, E. (2010). Promotion of self-regulated learning in classrooms: Investigating frequency, quality, and consequences for student performance. *Metacognition and Learning,* *5*(2), 157–171. https://doi.org/10.1007/s11409-010-9055-3

Molenaar, I., Horvers, A., & Baker, R. S. (2021). What can moment-by-moment learning curves tell about students' self-regulated learning? *Learning and Instruction, 72*. https://doi.org/10.1016/j.learninstruc.2019.05.003

Wiedbusch, M. D., Kite, V., Yang, X., Park, S., Chi, M., Taub, M., & Azevedo, R. (2021). A theoretical and evidence-based conceptual design of MetaDash: An intelligent teacher dashboard to support teachers' decision making and students' self-regulated learning. *Frontiers in Education*, 6. https://doi.org/10.3389/feduc.2021.570229

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In *Metacognition in educational theory and practice* (pp. 277–304). Lawrence Erlbaum Associates Publishers.

Van Leeuwen, A. (2015). Learning analytics to support teachers during synchronous CSCL: balancing between overview and overload. *Journal of Learning Analytics, 2*(2), 138–162. https://doi.org/10.18608/jla.2015.22.11

Verbert, K., Ochoa, X., De Croon, R., Dourado, R. A., & De Laet, T. (2020). Learning analytics dashboards: The past, the present and the future. In *ACM International Conference Proceeding Series* (pp. 35–40). Association for Computing Machinery. https://doi.org/10.1145/3375462.3375504

# Exploring Self-Regulated Learning Among Low-Achieving Students Using the Taiwan Adaptive Learning Platform

**Bor-Chen Kuo*[1]; Frederic Tao Yi Chang*[2]; Chih-Wei Yang*[3]**
National Taichung University of Education*
kbc@mail.ntcu.edu.tw[1]
snoopyfred@kimo.com[2]
yangcw@mail.ntcu.edu.tw[3]

**ABSTRACT**: Our study leverages educational big data and learning analytics to examine the self-regulated learning (SRL) of low-achieving students using the Taiwan Adaptive Learning Platform (TALP). We also investigate how factors like TALP engagement (measured in usage hours) and academic achievement influence SRL. The study involved 18,498 low achieving students from grades 3 to 8. SRL was assessed using the Self-regulated Learning Integrated Questionnaire (SRLIQ). Academic achievement was gauged through the Priori-tbt test, a standard assessment conducted by the Ministry of Education in Taiwan. Data on SRL, academic achievement, and TALP engagement were mined and extracted from the TALP database for analytical purposes. Our key findings include: (1) Higher grade levels correspond to lower levels of self-regulation, and females exhibit better self-regulation than males; (2) The correlation between SRL and academic achievement is approximately r=.23. In terms of TALP engagement, time spent on the platform after school shows a stronger correlation with SRL compared to time spent on instructional films, assessments, and logins; (3) The strength of the correlation between SRL and academic achievement varies across grades, with higher grades showing a stronger correlation. Finally, we sincerely acknowledge the National Science and Technology Council of Taiwan for their generous support of our project, MOST 109-2511-H-142 -002 -MY3.

**Keywords**: Self-regulated Learning, Academic achievement, Learning engagement, Online platform, TALP

## 1 INTRODUCTION

In the evolving field of education, Learning Analytics focused on Self-Regulated Learning (SRL) are crucial. These analytics, which involve measuring, collecting, analyzing, and reporting data, aim to enhance students' SRL experiences by aligning them with customized learning processes and environments (Long & Siemens, 2014). Modern digital learning platforms, such as the Taiwan Adaptive Learning Platform (TALP) endorsed by Taiwan's Ministry of Education, serve as vital repositories of educational data. These platforms, equipped with SRL questionnaires, effectively manage large-scale samples, facilitating a thorough investigation of SRL and the elements that influence it (Li, 2019). Numerous studies have shown a positive correlation between SRL and academic achievement, as well as its application in digital platforms (Dent & Koenka, 2016). However, limited research focuses on the SRL of low-achieving students, particularly in the K-12 demographic (DiFrancesca et al., 2016).

## 2. Methodology

In this study, we analyzed data from 18,498 students, identified as low-achieving through the Screen Test, a crucial component of the Project for Implementation of Remedial Instruction technology-based testing (Priori-tbt). The Priori-tbt, conducted by Taiwan's Ministry of Education, consists of two main assessments: the Screen Test, which filters low-achieving students, and the Progress Test, which evaluates the effectiveness of remedial instruction. These tests cover three subjects: Chinese, Mathematics, and English. Our sample, detailed in Table 1, included students from grades 3 to 8 participating in the TALP. The group consisted of 9,793 male and 8,705 female students. To assess their Self-Regulated Learning (SRL), our research employed the Self-regulated Learning Integrated Questionnaire (SRLIQ), an online tool in TALP. Furthermore, we monitored student engagement on TALP, tracking metrics such as hours spent on instructional films and tests, login durations, and after-school usage. Additionally, the Progress Test of Priori-tbt, encompassing Chinese, Math, and English, was used to measure academic achievement and the success of remedial instruction.

**Table 1: The distribution of participants across grade.**

| Grade | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 | Total |
|---|---|---|---|---|---|---|---|
| Participants | 2435 | 3748 | 6384 | 3393 | 1432 | 1106 | 18498 |

## 3. Result

The average SRL ability of the 18,498 users of the TALP is 0.07, with a standard deviation (SD) of 1.418. This SRL score is calculated using the Rasch scale. An in-depth analysis, as illustrated in Figure 1, shows variations in SRL abilities across grades 3 to 8. The breakdown by grade is as follows: Grade 3 has an average SRL score of 0.305 (SD = 1.376), Grade 4 has 0.193 (SD = 1.357), Grade 5 has 0.151 (SD = 1.409), Grade 6 has 0.043 (SD = 1.433), Grade 7 has -0.096 (SD = 1.432), and Grade 8 has -0.266 (SD = 1.397). Regarding gender differences, the average SRL score for male users is -0.042 (SD = 1.456), in contrast to 0.212 (SD = 1.370) for female users.

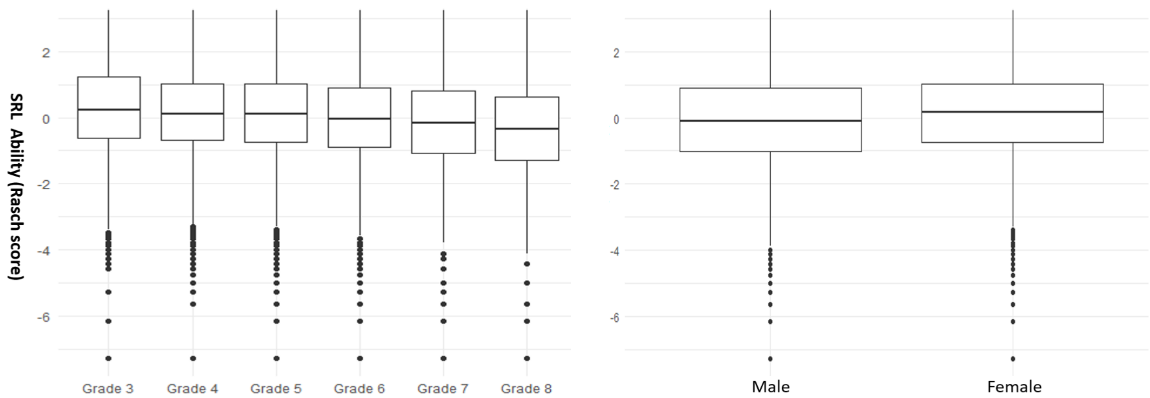



**Figure 1: The SRL ability across grade and gender**

As illustrated in Figure 2, the correlation between SRL and TALP engagement varies from .103 to .220. Similarly, the correlation between SRL and academic achievement ranges from .224 to .245. Regarding engagement with TALP, after-school use of the platform demonstrates the most significant correlation with SRL. Concerning the relationship between SRL and academic achievement, there is minimal variation in the strength of correlation across different subjects.

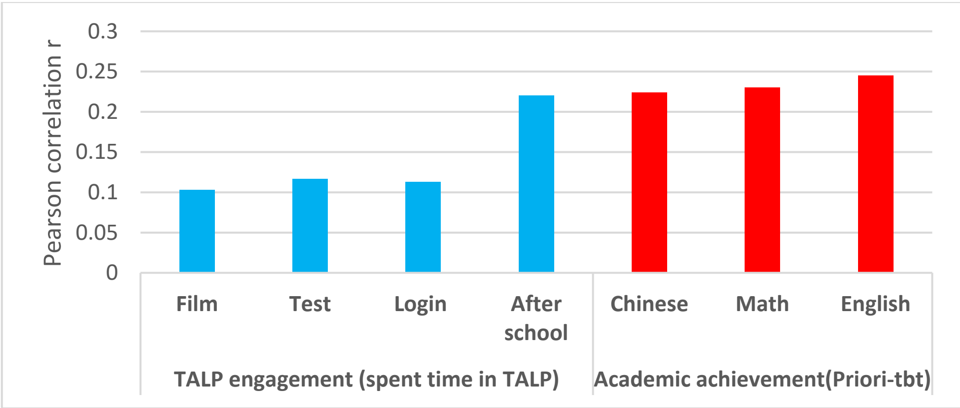



**Figure 2: The correlation of SRL between TALP engagement and academic achievement**

Figure 3 displays the range of correlation between Self-Regulated Learning (SRL) and academic achievement, which spans from .152 to .312. The heatmap indicates that, regardless of the subject—be it English, Mathematics, or Chinese—the magnitude of correlation increases with higher grade levels.

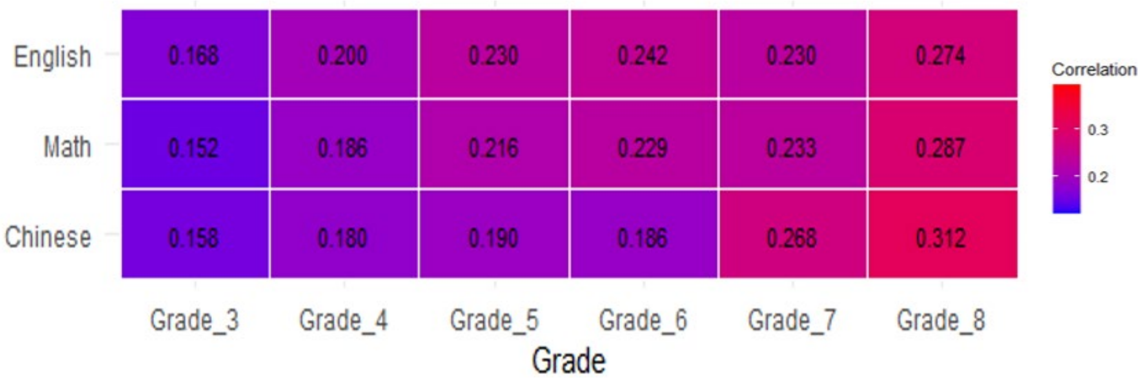



**Figure 3: The correlation between SRL and academic achievement across grades**

## REFERENCES


Long, P. & Siemens, G. (2014). Penetrating the fog: analytics in learning and education. *Italian Journal of Educational Technology, 22*(3), 132-137. https://www.learntechlib.org/p/183382/.

Schellings, G., & Van Hout-Wolters, B. (2011). Measuring strategy use with self-report instruments: theoretical and empirical considerations. *Metacognition and Learning, 6*(2), 83-90. https://doi.org/10.1007/s11409-011-9081-9

Li, K. (2019). MOOC learners' demographics, self-regulated learning strategy, perceived learning and satisfaction: A structural equation modeling approach. *Computers & Education, 132*, 16-30. https://doi.org/https://doi.org/10.1016/j.compedu.2019.01.003

DiFrancesca, D., Nietfeld, J. L., & Cao, L. (2016). A comparison of high and low achieving students on self-regulated learning variables. Learning and Individual Differences, 45, 228-236. https://doi.org/https://doi.org/10.1016/j.lindif.2015.11.010

Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. Educational Psychology Review, 28, 425-474.

# The Generative Multimodal Analysis (GMA) Methodology for Studying Socially Shared Regulation in Collaborative Learning

**Ridwan Whitehead[1], Andy Nguyen[2], Sanna Järvelä[3]**
Learning and Educational Technology Lab (LET Lab), University of Oulu
ridwan.whitehead@oulu.fi[1]
andy.nguyen@oulu.fi[2]
sanna.jarvela@oulu.fi[3]

**ABSTRACT**: This poster introduces the Generative Multimodal Analysis (GMA) methodology, an approach that enhances the study of learning contexts through the utilization of generative artificial intelligence to analyze multimodal data. GMA enables the large-scale interpretation of complex non-verbal behaviors, such as posture, gesture, and gaze, within collaborative learning contexts, thereby extending the empirical evidencing of these phenomena. By leveraging GenAI models with vision capabilities, the GMA model semi-automates the process of analyzing group interactions, which traditionally required extensive effort. We illustrate the application of GMA through two examples: a top-down approach of categorizing postural states as an alternative to pose estimation, and a bottom-up approach for multimodal analysis of a 1-minute video of a collaborative learning group to study socially shared regulation of learning. Preliminary results show a promising shift of paradigm which allows learning science researchers to comprehensively study non-verbal interactions in learning contexts. We argue that the use of GMA has the potential to contribute to collaborative learning analytics by offering a novel way to investigate and understand non-verbal behaviors.

**Keywords**: Socially shared regulation of learning, collaborative learning, generative AI

## 1 INTRODUCTION

Increasing evidence shows that regulation in group level contributes to successful collaborative learning (Haataja et al., 2022). Socially shared regulation in learning (SSRL) refers to a group's strategic, and transactive planning, task enactment, reflection, and adaptation. It involves groups taking metacognitive control of the task together through negotiated, iterative fine-tuning of cognitive, behavioral, motivational, and emotional conditions as needed (Järvelä et al., 2018). While traditionally, SSRL research has relied often on speech data, this approach overlooks the substantial non-verbal components of interaction. For instance, the posture of learners alone can provide valuable insights into learners' attitudes and engagement (Radu et al., 2020). However, conventional qualitative video coding methods are labor-intensive and impractical for large-scale applications. Although recent advances have introduced AI vision models, such as OpenPose or Azure Kinect to streamline this process, these techniques still present significant barriers in terms of accessibility, complexity, and resource demands. Furthermore, while machine learning models offer some promise in pose estimation, their reliability and success rates vary, while still requiring refinement and interpretation by researchers. In response to these challenges, this paper introduces a novel method we name “Generative Multimodal Analysis”, currently utilizing OpenAI's GPT-4-1106-vision-preview

API, proposing a more accessible, resource-efficient, and context-adaptable methodology for conducting non-verbal analysis in real-world collaborative learning settings.

# 2 GENERATIVE MULTIMODAL ANALYSIS TO ENHANCE MULTIMODAL LEARNING ANALYTICS FOR STUDYING SSRL IN COLLABORATIVE LEARNING

The Generative Multimodal Analysis (GMA) methodology, integrating generative AI, represents a cutting-edge approach for multimodal learning analytics by analyzing complex group interactions in learning contexts. At its core, GMA leverages sophisticated generative AI vision models, which have the capability of analyzing and describing visuals, including snippets of learning contexts. This approach can generate detailed descriptions of scenes, capturing the nuanced interplay between learners, both in static frames and in a temporal context. The fluidity of the GMA allows for a tailored analysis that adapts to the unique requirements of each dataset and context.

# 3 ILLUSTRATIVE CASE STUDIES

## 1.1 Case Study 1

Our first case spotlights how GMA can be used as an alternative to the manual labelling of posture and pose estimation by OpenPose or Azure Kinect. The SHARP dataset was used (see Järvelä & Nguyen, SHARP), which included a “make a breakfast smoothie” task that was captured using multimodal data collection (video, audio, etc.). To prepare the data for GMA, we automatically extracted the frames from the video recordings and sliced each frame into three equal parts, in which each participant was visible on their own. Each extracted image was looped in the model using our in-house prompts to obtain the participants’ postures as categories (e.g. leaning forward, backwards). Random sampling was conducted to measure the accuracy of the GenAI model, which achieved 81% accuracy in labelling the postures (Cohen’s kappa = 0.268). Stratified testing is required for improved assessment.

## 1.2 Case study 2

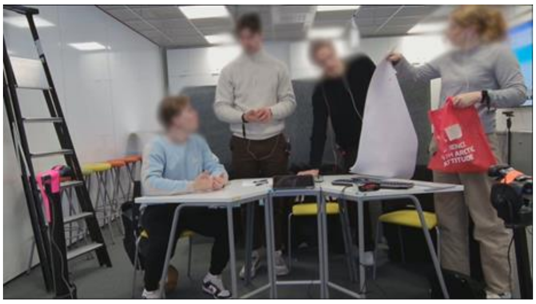


**Figure 1. A single frame of the video fed to the GenAI model**

Our second case study demonstrates the use of the GMA method to semi-automate multimodal analysis to study SSRL in collaborative learning. The data used is a one-minute-long video from a task on energy conversion, which involved a hands-on task (See Figure 1). Traditionally, it would have been cumbersome to analyze for multimodal interactions as it requires meticulous attention to detail and the ability to interpret multiple communication channels simultaneously, which can be resource-intensive and time-consuming. By using GMA, it is possible to extract detailed analysis from short duration video segments to deliver a detailed analysis of the interactions between the learners. The following output is a result of a 54 second video being fed to a GenAI with vision capabilities using in-house prompting, which as an example, could be used to exemplify interactions in a group:

> "…As the interaction unfolds, a participant standing initiates a non-verbal cue by reaching out for materials from a bag, creating an inclusive environment by sharing resources pertinent to the task at hand. This act of distributing materials non-verbally invites collaboration and potentially signals a shift in the focus of their activity. Simultaneously, there seems to be an exchange of ideas as indicated by hand gestures and eye contact among the participants. One participant's pointing gesture could be a mode of directing attention or providing instruction related to the energy conversion task, possibly correlating with specific content on the screen outside of the frame. The group's dynamics adapt as the communicative environment evolves; the standing participants lean in, showing deeper engagement, while the seated participants continue providing non-verbal feedback through nods and gestures that imply agreement, understanding, or encouragement…"

## 4 DISCUSSION AND CONCLUSION

The GMA methodology enhances SSRL research by allowing extensive analysis of multimodal data. GenAI interprets complex non-verbal behaviors in complex learning context, offering insights into the impact of these cues in SSRL. GMA semi-automates traditionally labor-intensive analysis, saving time and resources in studying collaborative learning interactions. It enables a deeper understanding of the dynamics present (Järvelä et al., 2018), improving collaborative learning design and support. This approach significantly contributes to multimodal collaborative analytics by facilitating the easy acquisition and integration of multimodal data streams, enriching the scope and depth of analysis (Järvelä et al., 2019). As GMA is in it's infancy, further testing with larger datasets and varied conditions, along with formal comparisons to other vision models and strategies for integrating its output with other tools, is essential for advancing the robustness and validity of this methodology.

## REFERENCES

Haataja, E., Dindar, M., Malmberg, J., & Järvelä, S. (2022). Individuals in a group: Metacognitive and regulatory predictors of learning achievement in collaborative learning. *Learning and Individual Differences*, *96*, 102146.

Järvelä, S., & Nguyen, A. SHARP - Socially Shared Regulation of Complex Learning Process in Groups (Version 1). University of Oulu, Learning & Educational Technology Research Unit (LET). https://doi.org/10.23729/ba2b81b0-580c-4f92-8e73-0659a514708c

Järvelä, S., Hadwin, A. F., Malmberg, J., & Miller, M. (2018). Contemporary perspectives of regulated learning in collaboration. In F. Fischer, C. E. Hmelo-Silver, P. Reimann, & S. R. Goldman (Eds.), Handbook of the learning sciences (pp. 127–136). Routledge.

Radu, I., Tu, E., & Schneider, B. (2020). Relationships between body postures and collaborative learning states in an augmented reality study. In *Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part II 21* (pp. 257-262). Springer International Publishing.

Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. A. (2021). What multimodal data can tell us about the students' regulation of their learning process?. *Learning and Instruction*, *72*, 101203.

# Exploring Human-AI Regulation in Design Contexts

**Ha Nguyen**
Utah State University
ha.nguyen@usu.edu

**Andy Nguyen**
University of Oulu
andy.nguyen@oulu.fi

**ABSTRACT**: Generative Artificial Intelligence (AI) tools present opportunities to facilitate creative ideation with human designers. In this work, we present case studies of human designers brainstorming ideas with ChatGPT, a generative AI tool, in user experience and user interface design contexts. We attend to how designers engage in **reflective design practices** to name tasks, frame the focus of their design, move to ideate, and reflect on design decisions. We also explore how designers approach the AI tool through **self-regulated learning** (SRL) activities to define and plan for the tasks, enact tactics, and evaluate the AI's responses to reframe the design steps. Analyses leverage Ordered Network Analysis to examine how reflective design practices and SRL co-occur in human-AI brainstorming and illustrate multifaceted learning engagement. Findings show how designers engage with novel, generative AI technology and illustrate the importance of SRL in AI-integrated learning.

**Keywords**: generative artificial intelligence, ordered network analysis, self-regulated learning

## 1 INTRODUCTION

Emergent research has positioned generative AI as a collaborative ideator that contributes novel insights in design (Karimi et al., 2020). In this paper, we explore how designers collaborated with ChatGPT (Chat Generative Pre-trained Transformer), a generative AI tool, in the context of user experience and user interface (UX/UI) design. To understand the AI-integrated brainstorming process, we examine how designers engage in *reflective design practices* to consider past actions and generate new insights (Schön, 1984). These practices comprise *naming* task aspects, *framing* design focuses, *moving* to brainstorm ideas, and *reflecting* on past decisions to reframe the design (Adams et al., 2003; Valkenburg & Dorst, 1998). In integrating AI, designers also engage in *self-regulated learning* (SRL) to *define* tasks, *plan*, *enact strategies*, and *evaluate* AI's responses (Järvelä et al., 2023). We ask: **How do designers engage in reflective design practices and SRL in collaborating with ChatGPT?** To answer this question, we applied Ordered Network Analysis (Tan et al., 2022), to examine the co-occurrences between reflective design practices and SRL as indicators of multifaceted learning engagement.

## 2 METHODS

### 2.1 Study setting and data sources

We drew from audio and screen-recorded interviews with 17 designers with varied backgrounds: six undergraduate, four graduate students (master's; PhD), and seven alumni from a Design program in a public university in the Mountain West United States. During their 45-minute, individual interviews, participants first brainstormed independently (without AI) to sketch ideas for a task (10 minutes):

Redesign the navigation of a learning management system. Participants refined their ideas with ChatGPT (15 minutes), and then sketched their designs in 5-7 minutes. Finally, they debriefed with the interviewers about their design ideas and the interactions with ChatGPT.

We focused on the brainstorming sessions with ChatGPT. We considered participants' think-aloud utterances at the sentence level as the main analysis units. Each transcript generated 88.53 utterances on average (*SD* = 32.32) for the brainstorming sessions. We built on prior work to code for reflective practices (Valkenburg & Dorst, 1998) and SRL based on the COPES model (Winne & Hadwin, 1998). The authors coded two transcripts separately to refine the code definitions. We reached substantial inter-rater agreement on a third transcript (Cohen's *k* range .74-1). We used a case study approach of three undergraduate participants (Alex, John, and Maya; pseudonyms), as they had similar prior design experience (rising senior, with no external internship), but demonstrated distinct patterns of collaborating with the AI. Specifically, Alex and John showed creative uses of the AI and were able to build on the generated ideas in their design sketches. Meanwhile, Maya only built on the AI's suggestions marginally and did not really reach an elaborated design by the end of the interview.

## 2.2 Ordered network analysis (ONA)

We leveraged ONA (Tan et al., 2022) to conceptualize learning as networks of co-occurring design practices and SRL strategies. ONA extends approaches like Epistemic Network Analysis (Shaffer et al., 2016) to emphasize temporal progressions. ONA visualizes the directions of co-occurrences when an activity more frequently precedes another and iterations of activities within an analysis window.

# 3 FINDINGS

## 3.1 Alex – flexible uses of AI for ideation

We present how Alex, John, and Maya approached ChatGPT differently. Alex's dominant reflective practices and SRL involved *naming* the tasks, followed by *planning* for the collaboration with the AI (arrows between "name" and "plan", panel A, Figure 1). For example, she began the session identifying the AI's gap and naming tasks (identifying student perspectives) in her think-aloud:

> **Alex [name-plan]:** Hmm! I think to be empathetic is human and I don't think there's enough opinions out there about Canvas use from a student's … So, I might start asking: What are the most important features for alerting management system to have when working with a K-12 audience?

After receiving the AI's response about features to include (e.g., user-friendly interface, course management, collaboration tools), Alex prompted the AI: "What are best practices for designing visually appealing and easy navigating websites for children?" Later on, she prompted ChatGPT to validate the AI (e.g., "cite sources"), ask for ideas (e.g., "give me ideas for an easier to navigate grade book"), make a template for comparing different systems, and role-play as an eight-year-old interacting with the tool. With each of these turns, Alex first *named* the tasks (e.g., user needs, product comparisons, ideation) in her think-aloud, before *planning* the approaches to using the AI tool.

Meanwhile, John's session was marked by iterative combinations of *planning*, *evaluating*, and making design *moves* and *tactics* (panel B; Figure 1). He gave clear role-play definitions for the AI in the beginning of his prompts, "Pretend that you are a UX/UI designer with 10+ years of experience", and then asked the AI to walk him through each step of the design process. After each step, John evaluated the extent to which the AI sufficiently responded to his prompts. When ChatGPT stated: "As a text-based AI, I'm unable to provide visual mockups directly", John stopped the response generation midway to rearticulate his prompt (i.e., *tactics*; changing "mockups" to "outline").

Finally, Maya's dominant strategies were *moving* and *tactics* (dark gray arrows between move and tactics, panel C, Figure 1). Most of her prompts (e.g., "What are the common stuff in an online course?") were to gather general information about the design space, with less iterations of planning and evaluation, compared to the other two participants.

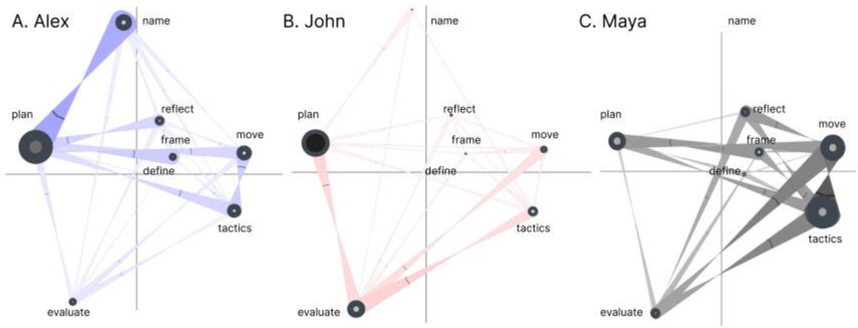



**Figure 1: Ordered Networks of Alex (A), John (B), and Maya (C)**

## 4 CONCLUSION

Our findings illustrate different approaches to brainstorming with generative AI, specifically highlighting the role of planning and evaluation, in conjunction with making design moves. We demonstrate the use of ONA in analyzing the link between SRL and design practices. We call for future work to investigate the importance of domain practices and SRL in AI-integrated learning.

## REFERENCES

Adams, R. S., Turns, J., & Atman, C. J. (2003). Educating effective engineering designers: The role of reflective practice. *Design studies*, *24*(3), 275-294.

Järvelä, S., Nguyen, A., & Hadwin, A. (2023). Human and artificial intelligence collaboration for socially shared regulation in learning. *British Journal of Educational Technology*.

Karimi, P., Rezwana, J., Siddiqui, S., Maher, M. L., & Dehbozorgi, N. (2020, March). Creative sketching partner: an analysis of human-AI co-creativity. In *Proceedings of the 25th International Conference on Intelligent User Interfaces* (pp. 221-230).

Schön, D. A. (1984). The architectural studio as an exemplar of education for reflection-in-action. *Journal of architectural education*, *38*(1), 2-9.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, *3*(3), 9-45.

Tan, Y., Ruis, A. R., Marquart, C., Cai, Z., Knowles, M. A., & Shaffer, D. W. (2022, October). Ordered network analysis. In *International Conference on Quantitative Ethnography* (pp. 101-116). Cham: Springer Nature Switzerland.

Valkenburg, R., & Dorst, K. (1998). The reflective practice of design teams. *Design studies*, *19*(3), 249-271.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In, DJ Hacker, J. Dunlosky, & AC Graesser. *Metacognition in educational theory and practice*, 277-304.

# A Social Network Analysis on Peer Connections in Leadership Development

**Author(s): Please Leave This Section Blank for Review**
Institution
Email

**Author(s): Please Leave This Section Blank for Review**
Institution
Email

**ABSTRACT**: The Director development Experience (DDE) is a leadership programme offered by the Civil Service College for first-time Directors in the Singapore Public Service. The DDE equips participants with the necessary perspectives, skills and support they require in their role. Evidence to gauge the effectiveness of the programme in developing peer connections as a form of professional support has been lacking. To address this, a study employing Social Network Analysis methods was conducted to measure the 2nd DDE's impact on fostering connections among participants. By gathering feedback data on networks based on desire to connect for work-related advice and to get-to-know-better, we harvested insights that implied the DDE was successful in establishing peer connections among participants.

**Keywords**: Singapore Public Service, Leadership Development, Adult Learning, Learning Design, Learning Groups, Social Network Analysis, Network Density, Community Detection, Centrality, Network Graph

## INTRODUCTION

The Director's Developmental Experience (DDE) at the Civil Service College Singapore is a leadership development programme for first-time directors in the Singapore Public Service; designed to equip directors with essential skills, perspectives, and support networks needed to perform effectively in their roles. The DDE Experience features an element called 'Learning Groups' which facilitate participants' learning experience in smaller and more intimate settings. Participants are assigned to learning groups based on criteria such as career experiences, job scope and gender to ensure a good mix of members. Despite the high value attached to the development of peer connections for professional support, there is currently insufficient evidence to demonstrate the programme's impact on developing the connections. This study used Social Network Analysis (SNA) methods to understand how peer connections were formed in the programme, to assess existing learning design (e.g. Learning Groups) and provide improvements for future iterations.

## METHODOLOGY

At the end of the 2nd DDE, we asked the participants to respond to two questions. The first question was related to ***work-related advice*** *("For each DDE participant in the table below, please indicate how likely you would approach him/her for work-related advice now that you have attended the DDE."). The second question was related to* ***get-to-know-better*** *("Who might you wish to get to know better* *after the programme?")*. We obtained 43 responses. For the "work-related advice" question, participants rated any number of their peers based on a scale ranging from "NA" (as unlikely as before the programme to reach out / Don't know the person / This is me) to "5" (definitely will reach out for work-related advice). Network graphs were constructed using participants as nodes and the ratings as weighted, bi-directed edges. The network density was utilised to assess the programme's effectiveness in establishing peer connections for professional support. Additionally, community detection was performed using machine learning algorithms on networks with mutually highly rated connections (i.e. rating 4 and above) to identify sub-groups. For the "get-to-know-better" question, participants rated any number of peers they wished to know better after the programme. Network graphs were created with participants as nodes and ratings as non-weighted, directed edges. We examined whether high in-degree centrality was linked to specific participant characteristics. Network graphs based on mutual ratings provided insights into participants' networking preferences, such as their desire to connect with a more diverse mix of peers outside of their learning groups.

## RESULTS AND FINDINGS

At the end of the programme, the "work-related advice" network achieved a network density of at least 73%, indicating a relatively high proportion of all possible connections were realised between the participants who were likely to approach each other for work-related advice and support. The community detection on the highly rated connections (rating 4 and above) revealed that Learning Groups played a significant role in forming subgroups, as shown in Table 1. This suggested that the design of Learning Groups could influence participants' connections with their peers.

**Table 1: Partial results from community detection showed most sub-groups consist of participants from the same learning group (LG).**

| Community A | |
|---|---|
| **Participant** | **LG** |
| Participant_1 | 9 |
| Participant_2 | 9 |
| Participant_4 | 9 |
| Participant_38 | 9 |
| **Community C** | |
| Participant_14 | 3 |
| Participant_26 | 3 |
| Participant_44 | 3 |
| Participant_20 | 4 |
| Participant_25 | 4 |
| Participant_35 | 4 |
| Participant_37 | 4 |

| Community B | |
|---|---|
| **Participant** | **LG** |
| Participant_5 | 1 |
| Participant_7 | 1 |
| Participant_10 | 1 |
| Participant_28 | 1 |
| Participant_39 | 1 |
| Participant_24 | 7 |
| Participant_27 | 7 |
| **Community D** | |
| Participant_8 | 2 |
| Participant_19 | 5 |
| Participant_22 | 5 |
| Participant_30 | 5 |

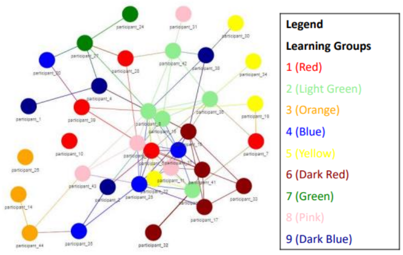
Participants with high in-degree centrality in the "get-to-know-better" network were also observed to be the more expressive participants who also won peer-based awards at the end of the DDE, as shown in Table 2. This suggests that specific characteristics or skill sets may influence network formation.

**Table 2: Participants with high in-degree centrality in "get-to-know-better" network and their characteristics based on remarks from programme owners.**

| S/N | Participant | In-degree Centrality (i.e. how many peers would like to know them better) | Remarks |
|---|---|---|---|
| 1 | Participant_38 | 21 | Expressive during the programme. Participant IC for one of the sessions. DDE award winner. |
| 2 | Participant_33 | 20 | Responsive to fellow participants' queries in WA chat group. DDE award winner. |
| 3 | Participant_42 | 19 | Participant IC for one of the sessions. DDE award winner. |
| 4 | Participant_14 | 17 | Expressive during the programme. Participant IC for one of the session. DDE award winner. |

The "get-to-know-better" network revealed that participants generally sought to connect with a diverse mix of peers beyond their learning groups, as depicted in Figure 1. This insight can guide decisions on enabling more cross-learning group interactions in future iterations.

**Figure 1: "Get-to-know-better" network showed that participants generally hope to know more diverse mix of peers beyond their learning group.**




## CONCLUSION, LIMITATIONS AND FUTURE WORK

The study showed that connection across Learning Groups could be improved, and this can be considered for future programme design. Other opportunities to collect data during the programme such as the mid-point could present new insights. However, data collection can be challenging especially as the programme continually seeks feedback data. Participants may experience feedback fatigue and decline participation. To address this, exploring trace data from social collaboration or communication platforms could be explored.

## REFERENCES

*Sarpy, S. A. C., & Stachowski, A. (2020). Evaluating Leadership Development of Executive Peer Networks Using Social Network Analysis. Journal of Leadership Education, 19(2)*

*Kennedy, J. C., Chan, L. K. Y., Chua, P. T., & Goh, L. T. (2005). Using social network analysis of peer ratings for leadership development. INTERNATIONAL MILITARY TESTING ASSOCIATION IN SINGAPORE, 47, 8-10.*

*Hoppe, B., & Reinelt, C. (2010). Social network analysis and the evaluation of leadership networks. The Leadership Quarterly, 21(4), 600-619.*

*Bartol, Kathryn & Zhang, Xiaomeng. (2007). Networks and Leadership Development: Building Linkages for Capacity Acquisition and Capital Accrual.*

*Carter, Dorothy & Dechurch, Leslie & Braun, Michael & Contractor, Noshir. (2015). Social Network Approaches to Leadership: An Integrative Conceptual Review. The Journal of applied psychology.*

# Understanding tutoring strategies of peer and expert tutors in online math discussions using Ordered Network Analysis (ONA)

**Yukyeong Song**
University of Florida
y.song1@ufl.edu

**Yingbo Ma**
University of Florida
yingbo.ma@medicine.ufl.edu

**Wangda Zhu**
University of Florida
wangdazhu@ufl.edu

**Wanli Xing**
University of Florida
wanli.xing@coe.ufl.edu

**ABSTRACT**: This poster aims to investigate the tutoring strategies that were used by middle school peer tutors and adult expert tutors while they help middle school students with mathematical problem-solving processes in an online asynchronous discussion board. To do this, we derived a theoretically grounded coding scheme for tutoring strategies and annotated the 20,776 discussion threads (manually and automatically). Then, we visualized the sequential dynamics of the tutoring strategies using the Ordered Network Analysis (ONA) method. Our preliminary results suggest that peer tutors use more direct guidance, such as giving answers, while expert tutors use more affective and metacognitive support. This study can inform the design of effective peer tutoring in mathematical discussion contexts.

**Keywords**: Mathematical problem-solving, Tutoring, Peer tutoring, Ordered network analysis

## 1 INTRODUCTION & BACKGROUND

Asynchronous online discussion forums supporting learners to interact with peers and experts are critical in STEM education. Tutoring procedures are most effective when thoroughly scaffolded (Sharpley & Sharpley, 1981), since untrained tutoring behaviors tend to be primitive, characterized by infrequent correction of errors and inappropriate giving of positive feedback (Topping et al., 2017). Previous research on tutoring in discussion forums focused on statics of tutoring behaviors such as tutor roles (Cho & Tobias, 2016) or peer tutoring styles (Smet et al., 2010). However, few studies investigated the dynamics of tutoring behaviors, especially that compared the peers' and experts' tutoring behaviors. This study investigates tutoring behavior dynamics among expert and peer tutors.

## 2 METHODS

We used the discussion data from Math Nation, an online math learning platform for secondary school students. Expert tutors are paid professional teachers who answer students' questions on Math

Nation; peer tutors are voluntary students helping other students to solve math problems. Students' posts on Math Nation are answered by expert tutors, peer tutors, or both types. We retrieved 20,776 threads that had more than three replies from the thread initiator out of 316,352 threads from 2015-06-01 to 2022-03-01.

### 2.1.1 Coding Scheme

Theoretically grounded in the literature of scaffolding (Pol, Volman, & Beishuizen, 2010), facilitation (Hmelo-Silver & Barrows, 2006), and peer tutoring (Chi, Siler, Jeong, Yamauchi, & Hausmann, 2001), we developed a coding scheme to annotate tutoring behaviors in discussion threads (Table 1).

Table 1. Tutoring strategy coding scheme

| Construct | Code | Description | Examples |
|---|---|---|---|
| Cognitive Scaffolding | Feedback | Provide confirmatory and correcting feedback. | Yes, that is correct. Not quite. |
| | Instructing | Giving step-by-step directions to students | Let's take this step by step. |
| | Explaining | Explaining the related concepts and principles, and providing additional information. | Don't get confused with the different variables. They all mean the same. |
| | Questioning | Asking questions to check their understanding | What is the value of the y-intercept? |
| Affective Support | Praising and encouraging | Giving praise for tutees' success and encouraging them to keep up their work. | That is correct! Way to go!!! :) |
| Metacognitive Support | Managing discussions | Managing the logistics of discussion and adjusting questions and answers, | Can you start a new thread? |
| Direct guidance | Giving answers | Directly giving answers to the question | The answer is C. |
| Tutoring Intervention | Encouraging peer tutoring | Encouraging (other) students' peer tutoring interactions | [Peer tutor Name], great helping. |
| | Guiding peer tutoring | Giving feedback on (other) students' peer tutoring interactions | Please make sure that we aren't posting any answers. |

### 2.1.2 Data Annotation

Two Ph.D. students manually annotated a partial dataset until they reached the desirable IRR (Cohen's Kappa = 0.822). Then, we trained automatic text classification models by fine-tuning RoBERTa, a pre-trained language model with state-of-the-art performance on a series of modeling text data in educational domains (Song et al., 2023) and collaborative learning dialogues (Ma et al., 2022). We used the Hugging Face xlm-roberta-based models to generate a 768-dimensional language embedding for each utterance. We split the annotated data into training (70%, n = 520) and evaluation (30%, n = 223) sets. The trained model achieved an accuracy of 0.75 and a macro F1 score of 0.73. Considering the number of labels (N= 8) is large, we deem this performance satisfactory. Lastly, we used the fined-tuned RoBERTa model to predict the rest of the dataset (n = 20,033).

## 2.2 Ordered Network Analysis (ONA)

Extended from Epistemic Network Analysis (ENA), ONA captures temporal connections of coded utterances and visualizes the frequency of each node and the strength and direction of the connections (Fan, et al., 2023). In our study, each tutor's utterances were used as *lines*, and each single discussion thread as *conversations.* The *units of analysis* in this study were mathematical discussion threads.

# 3 RESULTS

ONA results suggest that cognitive scaffolding (e.g., questioning, explaining) was the most outstanding strategy found in both peers and experts. However, peer tutors tended to give more direct guidance (e.g., giving answers) and instructions followed by explanations. Expert tutors tended to provide more affective support such as praising, and metacognitive support such as managing discussions.

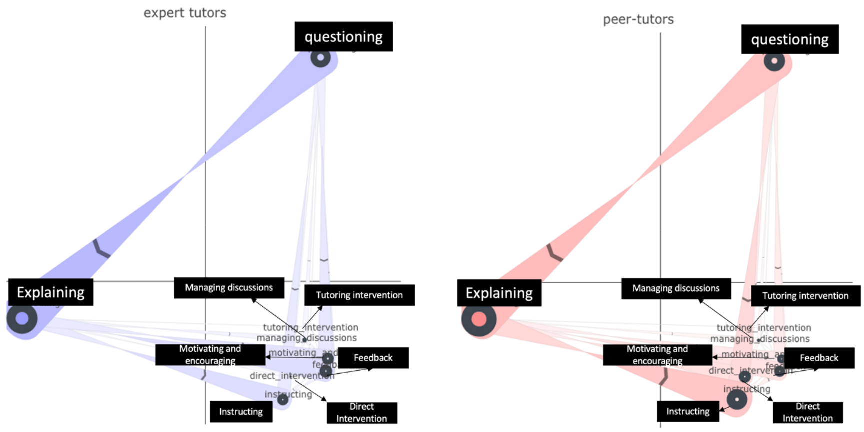



Figure 1. Ordered Network of Peer and Expert Tutors

# 4 DISCUSSION & CONCLUSION

We found similar patterns in both peer and expert tutors' use of tutoring strategies. However, some granular differences provide us with insights into how to design effective tutoring activities. Direct interventions by peer tutors may be helpful in certain learning contexts; however, it also raises concerns about the development of critical thinking skills among students when consistently provided with straightforward answers. The ONA highlighted the critical role of expert tutors in shaping the tutoring dynamics, as their interventions focused on providing metacognitive support and affective support. However, the expected role of "tutoring intervention" (e.g., guiding peer tutoring) was rarely found in our results. This could be because of the small percentages of this code in our dataset. In our future studies, we will consider combining similar codes based on their locations in the ONA plots.

# REFERENCES

Chi, M. T., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive science, 25*(4), 471-533.

Cho, M. H., & Tobias, S. (2016). Should instructors require discussion in online courses? Effects of online discussion on community of inquiry, learner time, satisfaction, and achievement. *International Review of Research in Open and Distributed Learning, 17*(2), 123-140.

Fan, Y., Tan, Y., Raković, M., Wang, Y., Cai, Z., Shaffer, D. W., & Gašević, D. (2023). Dissecting learning tactics in MOOC using ordered network analysis. *Journal of Computer Assisted Learning*, *39*(1), 154-166.

Sharpley, A.M. and Sharpley, C.F. (1981) Peer tutoring: A review of the literature. Collected Original Resources in Education, 5, 7-C11.

Hmelo-Silver, C. E., & Barrows, H. S. (2006). Goals and strategies of a problem-based learning facilitator. *Interdisciplinary journal of problem-based learning, 1*(1), 4.

Song, Y., Xing, W., Tian, X., & Li, C. (2023, March). Are We on the Same Page? Modeling Linguistic Synchrony and Math Literacy in Mathematical Discussions. In *LAK23: 13th International Learning Analytics and Knowledge Conference* (pp. 599-605).

Topping, K., Buchs, C., Duran, D., & Van Keer, H. (2017). *Effective peer learning: From principles to practical implementation*. Taylor & Francis.

Ma, Y., Celepkolu, M., & Boyer, K. E. (2022, March). Detecting impasse during collaborative problem solving with multimodal learning analytics. In *LAK22: 12th International Learning Analytics and Knowledge Conference* (pp. 45-55).

# Learning analytics using epistemic network analysis in a mathematical reasoning intelligent team tutoring system

**Bor-Chen Kuo**
National Taichung University of Education
kbc@mail.ntcu.edu.tw

**Chia-Hua Lin**
National Taichung University of Education
linch2812@gmail.com

**Shu-Chuan Shih**
National Taichung University of Education
ssc@mail.ntcu.edu.tw

**ABSTRACT**: This research presents a novel approach to enhancing mathematical reasoning skills in students through a Chinese dialogue-based, intelligent team tutoring system. This system facilitates collaborative learning by encouraging students to engage in mutual discussions and collectively reorganize problem-solving strategies. Our study employs epistemic network analysis (ENA) to investigate the dialogic interactions within the system, providing insights into the developmental trajectory of students' mathematical reasoning abilities. We specifically focus on comparing the progress and interaction patterns across different student groups. This exploration not only underscores the effectiveness of dialogue-based learning in mathematics but also contributes to a deeper understanding of how collaborative learning environments can be optimized for skill development in reasoning. Additionally, we use ENA to gain insights into the differences in dialogue processes between high and low-ability grouped students.

**Keywords**: learning analytics, epistemic network analysis, intelligent team tutoring system

## 1 INTRODUCTION

In Taiwan's educational landscape, traditional classroom teaching, predominantly lecture-based, limits student engagement and interaction (Chang et al., 2012). Recognizing the need for enhanced student involvement in mathematical learning, this study incorporates with student interaction model embed into the Intelligent Tutoring System (ITS) to facilitate mathematical dialogues and reasoning. While ITS has made strides in teaching fundamental mathematical concepts and problem-solving, its application in fostering mathematical practice abilities remains underexplored. This study, therefore, employs an innovative team-based ITS approach to augment students' mathematical reasoning skills. Additionally, it utilizes Epistemic Network Analysis (ENA) to analyze the dialogic interactions within this system, aiming to assess the development of students' mathematical reasoning abilities and the effectiveness of collaborative learning processes.

## 2 METHOD

This study employed the Chinese mathematical intelligent team tutoring system (ITTS), an online learning platform created based on the theoretical framework of AutoTutor (Nye et al., 2014) and the Chinese mathematical intelligent tutoring system (Kuo et al., 2019; Pai et al., 2021). The ITTS extends the standard ITS architecture with a team model alongside the conventional four-model framework (domain, tutoring, interface, and student models). It features an adaptive grouping mechanism that pairs students based on their abilities for collaborative online discussions.

The experiment involved 68 sixth-grade students from a Taiwanese elementary school. Within the ITTS environment focused on mathematical reasoning, students were paired—combining one high-ability with one low-ability student—based on ability assessments. Each pair worked collaboratively on seven tasks from the 'How to Solve Math Problems' unit.

In this study, we utilized epistemic network analysis (Bowman et al., 2021; Shaffer, 2017; Shaffer, Collier, & Ruis, 2016; Shaffer & Ruis, 2017) on our dataset through the ENA Web Tool (version 1.7.0; Marquart et al., 2021). Our ENA model included seven codes, based on the mathematical reasoning norms summarized by Lin (2012) (Table 1) and two additional discourse codes specific to this study: student-initiated teaching (ST) and discourse unrelated to reasoning norms (NS).

**Table 1: The mathematical reasoning norms (Quoting from Lin, 2012)**

| Codes | Reasoning norms |
|---|---|
| KC | To find out the reason from various mathematical symbols. |
| KR | To focus on the statement to explicate the solution strategies and what knowledge to use. |
| MA | Building the criteria of categorization. |
| MQ | Question and evidence. |
| MG | Deducing the regulations from examples. |

## 3 RESULTS

Our application of Epistemic Network Analysis (ENA) yielded network graph (Figures 1), which illustrate the distinct learning trajectories in mathematical reasoning skills among students. Figure 1 highlight the performance disparities in reasoning skills between high and low-ability groups.

Specifically, a two-sample t-test, assuming unequal variance along the X-axis, indicated significant differences between high-ability groups (mean=0.09, SD=0.19, N=34) and low-ability groups (mean=-0.09, SD=0.16, N=34; t(64.14)=4.10, p=0.00, Cohen's d=1.00) at the alpha=0.05 level. This indicates a pronounced distinction in the proficiency of mathematical reasoning skills.

As depicted in Figure 1, high-ability students exhibited a more advanced performance across most reasoning skills, notably engaging in proactive teaching behaviors (ST) during discussions. In contrast, low-ability students often engage in discussions that are irrelevant to reasoning (NS), transitioning from known information to unknown areas (KC, KR) and building the criteria of categorization (MA).

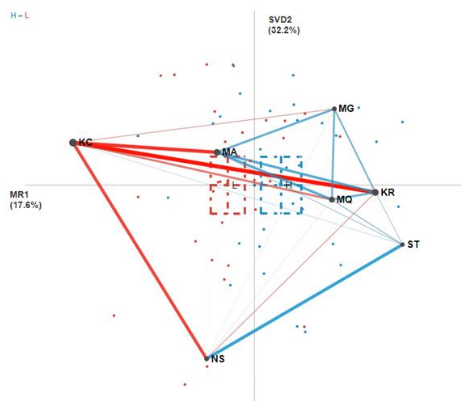



**Figure 1: Comparison network for high (blue) and low-ability (red) groups**

This study was funded by the National Science and Technology Council of Taiwan (Grant No. MOST 111-2410-H-142-031-MY3.)

## REFERENCES

Bowman, D., Swiecki, Z., Cai, Z., Wang, Y., Eagan, B., Linderoth, J., & Williamson Shaffer, D. (2021). The mathematical foundations of epistemic network analysis. In A. Ruis & S. B. Lee (Eds.), Advances in Quantitative Ethnography: Second International Conference, ICQE 2020, Malibu, CA, USA, February 1–3, 2021, Proceedings (pp. 91–105). Springer.

Chang, C. Y., Chen, C. H., Fun, D. S., & Lin, J.J. (2012). The Development of Formative Assessment Practice through Teachers' Professional Learning Community and Its Impact on Student Learning Achievement. *Bulletin of Educational Psychology, 43*(3), 717-733. https://doi.org/10.6251/BEP.20110318

Kuo, B. C., Lin, C. H., Pai, K. C., Shih, S. C., & Liao, C. H. (2019). Using learning analytics to explore the performance of Chinese mathematical intelligent tutoring system. In *Adaptive Instructional Systems: First International Conference, AIS 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings 21* (pp. 432-443). Springer International Publishing.

Marquart, C. L., Hinojosa, C., Swiecki, Z., Eagan, B., & Shaffer, D. W. (2021). Epistemic network analysis (Version 1.7.0) [Software]. Available from http://app.epistemicnetwork.org

Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education, 24*, 427-469.

Pai, K. C., Kuo, B. C., Liao, C. H., & Liu, Y. M. (2021). An application of Chinese dialogue-based intelligent tutoring system in remedial instruction for mathematics learning. *Educational Psychology*, *41*(2), 137-152.

Shaffer, D. W. (2017). Quantitative ethnography. Madison, WI: Cathcart Press.

Shaffer, D. W., & Ruis, A. R. (2017). Epistemic network analysis: A worked example of theory-based learning analytics. In C. Lang, G. Siemens, A. F. Wise, & D. Gasevic (Eds.), *Handbook of learning analytics* (pp. 175–187). Society for Learning Analytics Research.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics, 3*(3), 9–45.

# A Dynamic Time Warping Approach to Time-Series Clustering Based on Learners' Backtrack Reading Rates

**Minjung Kim & Hyun Sook Yi***
Konkuk University, South Korea
j145950@naver.com, hyunsyi@konkuk.ac.kr*

**ABSTRACT**: This study investigates learners' reading behaviors in online learning environments, with a specific emphasis on backtrack reading. Using a substantial dataset comprising 1,474,680 rows of page-stream data from an e-book reading program designed for elementary students, the analysis scrutinizes the learning logs of 203 students spanning multiple weeks. Dynamic time warping was employed to cluster students based on their daily reading logs. Following the exploration of the optimal number of clusters, two clusters exhibiting distinct patterns of temporal sequences were identified. It was observed that the number of words learned did not show a significant difference between the two groups. However, students with higher backtrack reading rates demonstrated notably lower pronunciation scores and persistence of reading compared to students with lower backtrack reading rates.

**Keywords**: time-series clustering, dynamic time warping, backtrack reading rate

## 1 INTRODUCTION

With the growing expansion of online learning in K-12 education, there has been a significant increase in research on learners' cognitive and affective characteristics through analyses of log data. Learning behaviors exhibited by students during on-line learning can be directly observed from records in log data, such as notes taken by learners or quiz results. Alternatively, researchers may formulate new variables by combining some observable variables to unveil hidden aspects of learner behaviors. Among various types of information extracted from log data, this study specifically focused on the backtrack reading, which provides insights into learners' reading behaviors (Yin et al, 2018). A time-series clustering analysis was conducted to comprehend students' reading patterns based on daily reading logs within an on-line reading program designed for elementary school students and performance scores and persistent reading habits were compared among the identified clusters. The research questions are as follows: (1) How many clusters can be identified from backtrack reading rates? (2) How do the clusters differ in terms of performance of reading and persistent reading habits?

## 2 METHOD

The data used in this study comprises 1,474,680 rows of page-stream data collected from May to June 2023 through an e-book reading program tailored for elementary school students. Our analysis focused on the learning logs of 203 students for whom reading records spanning more than one week were available. The e-book reading system encompasses 5 steps: Steps 1 and 2 involve vocabulary learning, Step 3 is dedicated to reading activities, and Steps 4 and 5 involve speaking programs with sentence reading tests. Before conducting the time-series cluster analysis, the backtrack reading rate was computed using the number of times a student turns to previous or subsequent pages. As another measure of reading behaviors, persistence of reading was derived by calculating the difference between the number of books started in Step 1 and the number of books completed in Step 3. As observable measures, the study used the number of words studied for each month, the amount of time spent on each Step, daily learning time, and pronunciation scores.

As a method for clustering students based on their daily reading logs, a dynamic time warping (DTW) was employed to measure the similarity between two sequences of learning behaviors. The DTW is particularly advantageous for clustering temporal log data with variations in length, duration, and timing of events (Sakoe & Chiba, 1978). Following the transformation of users' page-stream into daily reading process data, the partition around medoids (PAM) algorithm based on DTW distance was used to classify students with similar backtrack reading rates, as it provides accurate results with small samples (Kaufman & Rousseeuw, 1990). Data analysis and visualization were carried out using 'dtwclust' package in R (Sardá-Espinosa, 2019).

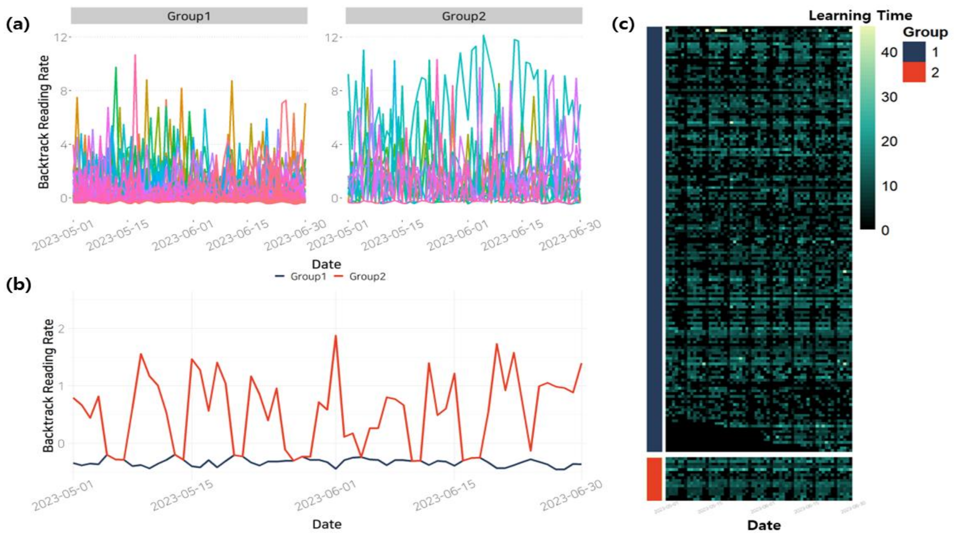



**Figure 1: Patterns of daily backtrack reading rate by cluster**

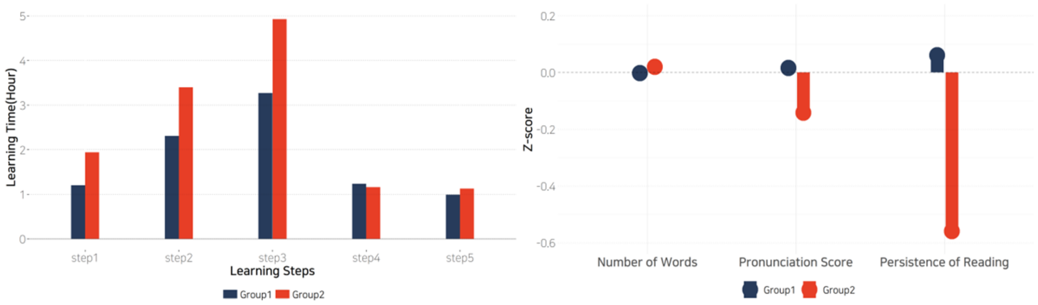



**Figure 2: Reading behaviors and learning outcomes by cluster**

## 3 RESULTS AND DISCUSSION

The exploration of the optimal number of clusters involved incrementing the number of clusters from 2 to 5 with 10 repetitions. The PAM clustering with two clusters exhibited the best CVI values (Silhouette: 0.687; Calinski-Harabasz: 64.617; Dunn: 0.101; Davies-Bouldin Index: 1.679; and COP: 0.145). The daily backtrack reading rates are visually depicted both at the individual student level

(Figure 1(a)) and at the group level using medians(Figure 1(b)). Group 1 exhibits higher backtrack reading rates compared to Group 2, although there is overlapping during certain time periods. To assess whether the observed backtrack reading rates might be attributed to low learning time during the overlapping periods in Figure 1(b), the daily learning time of individual students was rearranged by groups (Figure 1(c)). Since the difference in daily learning time between the two groups is not noticeable, the disparity in the trends of backtrack reading rates can be interpreted as a reflection of differences in reading behaviors rather than variations in learning time.

The left panel of Figure 2 illustrates the daily learning time of the two clusters on each step. Notably, Group 2 exhibits significantly higher daily learning time than Group 1, particularly on Step 1 through Step 3, with no significant differences observed on Steps 4 and 5. The right panel of Figure 2 presents the number of words learned, pronunciation scores, and persistence of reading of the two clusters. It is observed that the number of words learned does not exhibit a significant difference between the two groups. However, Group 2 demonstrates notably lower pronunciation scores and persistence of reading compared to Group 1. Part of the reasons for the lower performance of Group 2 on pronunciation scores could be attributed to lower learning time on Steps 4 and 5, which are designated for sentence speaking activities. Furthermore, students in Group 2 invested more time in Steps 1 to 3, where actual reading activities occur, but had lower reading persistence. This implies that students in this cluster may have difficulty completing reading of an entire book.

## LIMITATIONS AND FUTURE WORK

Defining hidden learning behaviors is not a straightforward task and analyzing them demands more intricate statistical modeling. In this study, the backtrack reading behaviors did not emerge as desirable characteristics. There is a need for further investigation about learners' reading behaviors along with other characteristics to gain a more comprehensive understanding of their behaviors.

## ACKNOWLEDGEMENT

This work was supported by Institute of Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No. RS-2023-00229780). The data were provided from the 1st Edutech Data Analysis Competition hosted by the Society for EduTech.

## REFERENCES

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in data: An introduction to cluster analysis*, John Wiley & Sons, New York.

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. doi: 10.1109/tassp.1978.1163055

Sardá-Espinosa, A. (2019). Time-Series Clustering in R Using the dtwclust Package, *The R Journal, 11*(1), 22-43. doi:10.32614/RJ-2019-023

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 Facor Approach. *International Journal of Human-Computer Interaction, 35*(4-5), 313-322. doi: 10.1080/10447318.2018.1543077

# AbstNebula: A Method to Visualize the Back-and-forth between Concrete and Abstract Utterances for Discussion Analysis

**Ryunosuke Nishimura, Chikae Watanabe, Hironori Egi**
The University of Electro-Communications
n2230106@edu.cc.uec.ac.jp, hiro.egi@uec.ac.jp

**ABSTRACT**: Visualizing discussions that involve back and forth between concrete and abstract utterances is useful for understanding the processes by which learners develop their ideas. This study introduces AbstNebula, a novel tool for visualizing the abstractness of utterances in discussions. It aims to describe how learners develop their ideas and provides information to reflect on the discussion. AbstNebula transcribes the utterances in discussions and calculates the nouns' abstractness on a scale from concrete to abstract. The results are then visualized as polar coordinates, where the radius indicates abstractness, and the angle denotes discussion time. Using this method, abstract words are displayed in the form of a nebula at the center of the polar coordinates, and concrete words are placed at the periphery. An experiment with participants discussing "What is thoughtful consideration?" revealed the potential of AbstNebula to visualize the development of participants' ideas to generalized abstract conclusions from concrete experiences. The findings of this study suggest that AbstNebula provides a visual framework to understand the progression of ideas, and it can effectively facilitate reflection on discussions. Future work will integrate AbstNebula into systems where reflection outcomes can be enhanced.

**Keywords**: Discussion Analysis, Visualization, Abstractness of Words, Reflection

## 1 INTRODUCTION

Visualization of discussions after they have taken place is a major approach to help learners deepen their thinking based on analysis of multimodal data. In particular, visualization of linguistic information, such as the content of utterances, reveals how the participants in a discussion develop their ideas objectively. Reflecting on discussions can be useful for such a style of learning. However, simple transcription and display of the discussion do not encourage deep reflection. This is because a deeper reflection requires an understanding of the context of the discussion (Moon, 2013). For example, the visualization of a speech using a word cloud may encourage critical reflection on the discussion (DeNoyelles et al., 2015). The word cloud identifies important utterances based on the frequency of the utterances and presents them visually. However, it does not capture the process by which learners develop their ideas over time. Therefore, it will remain up to the learners to generalize their experiences and refer to concrete examples in the discussion. In other words, they will be expected to deepen their ideas by climbing up and down "the ladder of abstraction" (Hayakawa et al., 1978). It is warranted that the ideas that learners develop during the discussion can be clearly described by identifying concrete and abstract utterances, followed by visualizing the back and forth between them.

In this study, we propose AbstNebula which visualizes the transition between the content of concrete and abstract utterances. We aim to describe how learners develop their ideas and provide useful information for reflection on the discussion.

## 2 METHODOLOGY

Figure 1 shows an overview of the mechanism of working of the AbstNebula system, along with a situation of some participants involved in a discussion. The system first transcribes the audio of the discussion. Next, it performs a morphological analysis of the transcribed text and extracts nouns from it. The system then calculates the abstractness of the extracted nouns and creates visuals that are shown to the participants in the discussion to help them reflect on the development of their ideas.

Figure 2 depicts an AbstNebula that visualizes the back and forth between concrete and abstract nouns uttered by a participant who spoke the highest number of nouns during the discussion; these nouns were classified based on their abstractness. Because the discussion was conducted by a Japanese speaker, the Word Database for Japanese common words (NAIST, 2019) was used to calculate the abstractness of words. In the database, the abstractness of every word is represented by a real number ranging from 1 to 5. The closer the abstractness is to 1, the more concrete the word is, and the closer it is to 5, the more abstract the word. For example, the abstractness of “Tokyo” is 1.3 and that of “city” is 2.0 in the database. The back and forth between concrete and abstract nouns for each participant was visualized using a polar coordinate, where the radius is 5 minus the abstractness value of the noun; the angle is associated with the time elapsed in the discussion. The angle starts from the north direction and increases clockwise as the discussion progresses. After it goes around the circle, it ends at the starting point. As a result, abstract nouns are displayed in the form of a nebula in the center of the polar coordinate, and concrete nouns are displayed toward the periphery of the polar coordinate. A given word is displayed only once to encourage reflection from a variety of perspectives.

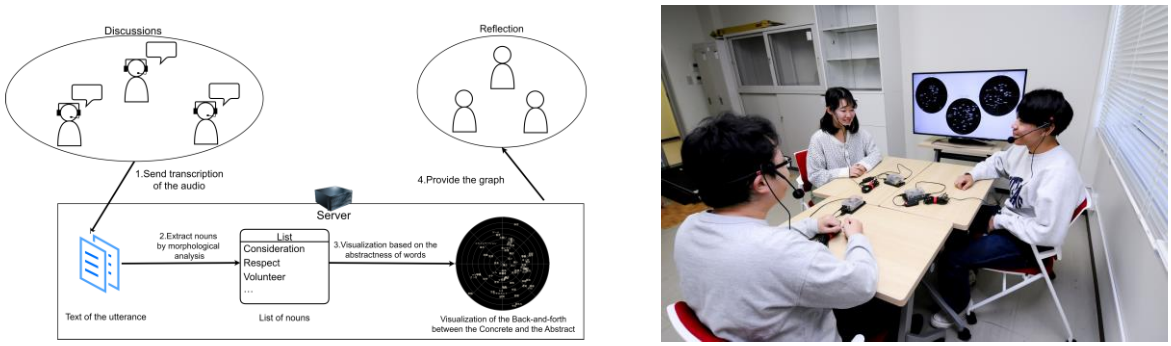



**Figure 1: Overview of the method (left) and situation of participants during the discussion (right)**

## 3 APPLICATION ON A DISCUSSION AND RESULTS

An experiment was conducted to evaluate the effectiveness of AbstNebula. Each participant took part in a discussion on “What is thoughtful consideration?”. These discussion sessions were attended by graduate and undergraduate students from a science and engineering university. The discussion had three participants in each group and lasted for 10 minutes. The system was applied to 9 groups.

Figure 2 shows an AbstNebula for utterances of one participant in a randomly selected group who uttered the highest number of nouns during a discussion on the aforementioned topic. At the periphery of the nebula, we can see that this participant shared giving up one’s seat to an “elder” on a "bus" as a concrete example of thoughtful consideration. In the central part of the nebula, we can see that this participant considered taking "action" have "importance" in thoughtful consideration. In addition, by reading the words uttered at ~2.5 minutes from the center to the periphery, this participant gave examples of thoughtful consideration based on "calculation," such as expecting something in "return", such as "money." The difference in morphology is due to the difference between Japanese and English.

In conclusion, AbstNebula enables the visualization of the back and forth between utterances based on concrete and abstract thinking. AbstNebula will be applied to a discussion reflection system.

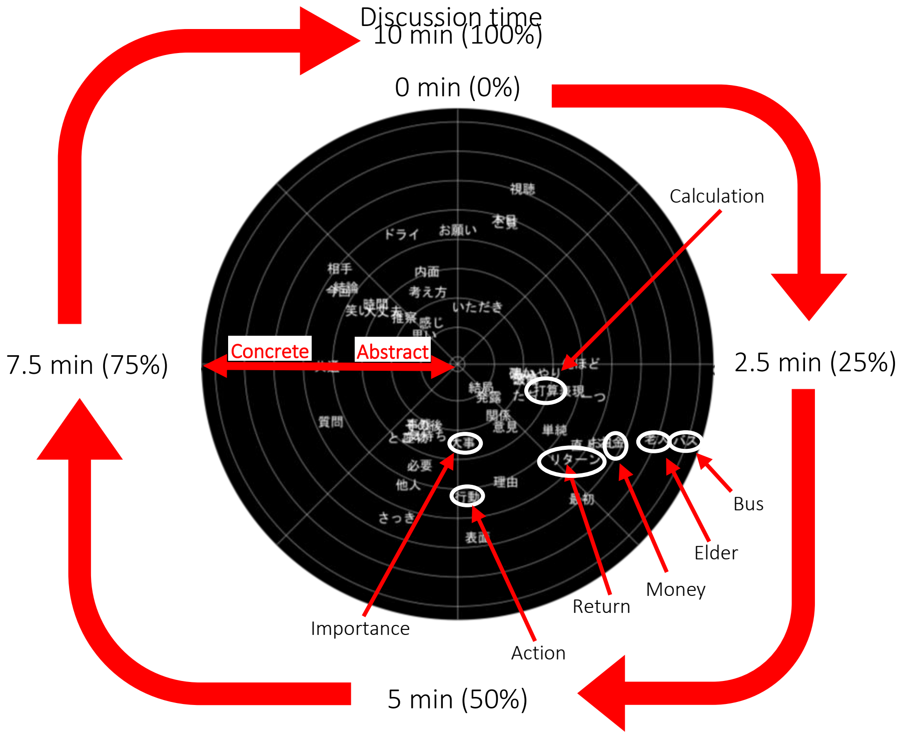



**Figure 2: An AbstNebula based on a discussion on "What is thoughtful consideration?"**


## ACKNOWLEDGEMENTS

This work was partly supported by the Grants-in-Aid for Scientific Research (NOs. 21K02752, 21K18527 and 22K02951) by MEXT (Ministry of Education, Culture, Sports, Science and Technology) in Japan.


## REFERENCES


DeNoyelles, A., & Reyes-Foster, B. (2015). Using word clouds in online discussions to support critical thinking and engagement. *Online Learning*, *19*(4), 13-24. https://doi.org/10.24059/olj.v19i4.528

Hayakawa, S. I., & Hayakawa, A. R. (1978). *Language in thought and action* (4th ed.). New York: Harcourt Brace Jovanovich.

Moon, J. A. (2013). *A handbook of reflective and experiential learning: Theory and practice*. Routledge.

NAIST, Social Computing Laboratory. (2019) AWD-J: Abstractness of word database for Japanese common words. https://sociocom.naist.jp/awd-j/

# An Approach of Multimodal Learning Analytics based on the Distance between Learners' Heads during Collaborative Learning

**Sei Yamamoto[1], Ryo Funabashi[1], Takeo Noda[2], Masataka Kaneko[2], and Hironori Egi[1]**
[1]The University of Electro-Communications, [2]Toho University
y2010713@edu.cc.uec.ac.jp, hiro.egi@uec.ac.jp

**ABSTRACT**: Typically, evaluating collaborative learning requires expert observation and analysis. Previous studies have suggested the potential of measuring various body movements for evaluation. Thus, this paper proposes a method to evaluate collaborative learning by measuring the distance between learners' heads continuously using a depth camera mounted directly above the learners. This distance is hypothesized to reduce when learners are more engaged in the learning topic. The proposed method is intended to reduce the influence of measurement procedures on learners in actual learning environments. The proposed method can be used by teachers to determine whether learning is progressing as intended. The results of a preliminary experiment conducted with two learners and a tablet device with an interactive math material indicate that the proposed method can measure the distance between learners' heads automatically and continuously. In the future, we plan to evaluate the distance between learners' heads against expert evaluations and the learners' internal states. In addition, we plan to integrate this distance information with other multimodal information to further improve the proposed method.

**Keywords**: Distance between learners' heads, Collaborative learning, Multimodal learning analytics

## 1 INTRODUCTION

Evaluating collaborative learning requires focusing on both the outcome and the learning process. This process also requires continuous observation by experts of the learning topics. Extracting evaluation indices is dependent on the expert's ability; thus, reproducibility cannot be guaranteed for evaluations that require impartiality. In this study, we attempt to analyze learning objectively by measuring the distance between learners' heads. The proposed method is expected to provide practical benefits to multimodal learning analytics.

A previous study investigated the features among group work behaviors, which are success factors in open-ended tasks, by acquiring multimodal information about the learners in group work and using machine learning technology (Spikol et al., 2018). Results demonstrated that the distance between the learners' hands and faces can be used to estimate learning activities' artifact quality. Another study collected multimodal data from video data during collaborative problem-solving (CPS) and constructed a decision tree to estimate whether the learners perform the actions observed in the CPS process (Cukurova et al., 2020). The results demonstrated that it is possible to determine learner behavior from video data, e.g., the distance between a learner's left and right hands, the distance between learners' bodies or faces, and the number of learners' face in-the-screen. These studies have confirmed that estimating the learning situation by measuring learners' body movements is possible.

Another previous study defined the learner's motivation level in the computer-supported collaborative learning context using a tablet device (Funabashi et al., 2022). Here, the learner's motivation level was estimated by acquiring information on where their hands were positioned in the area around the desk and tablet device. However, using the previously proposed method, it is difficult to identify groups that fail in collaboration because the level of motivation can only be calculated for individual learners and does not focus on more complex relationships among multiple learners.

## 2 PROPOSED METHODOLOGY

We hypothesize that learners may move closer to each other during collaborative learning processes when they are interested in the learning topic. This study focuses on continuous measurement of the distance between learners' heads. Here, we assume that the information used to measure the distance between learners' heads is obtained from above the learners. Compared with the method that measures from in front of the learners, measuring from above has two distinct advantages. First, it is easy to obtain only the target learners; thus, the proposed method can be employed to analyze collaborative learning involving a large number of learners. Second, the measurement equipment required to measure from above is less visible to the learners. Easily visible measurement equipment may cause a loss of concentration due to the thought of being monitored. Measuring from above, e.g., from the ceiling, can be implemented easily in practical learning environments, and its influence on learning can be minimized because performing such measurements from above makes it easy to identify the target learners.

Therefore, we propose a method that measures the distance between learners' heads from above. This allows teachers to analyze learning activities in collaborative learning scenarios easily by presenting the distance between the learners' heads. Based on our hypothesis, the teacher can understand when learners become more interested in the topics during the learning activities. In addition, with the proposed method, teacher burden can be reduced because the proposed method does not require watching the entire recorded video to identified significant points. The proposed method is also useful for teachers in terms of reflecting on whether learning is progressing as intended and whether the teaching materials are appropriate for the given learners.

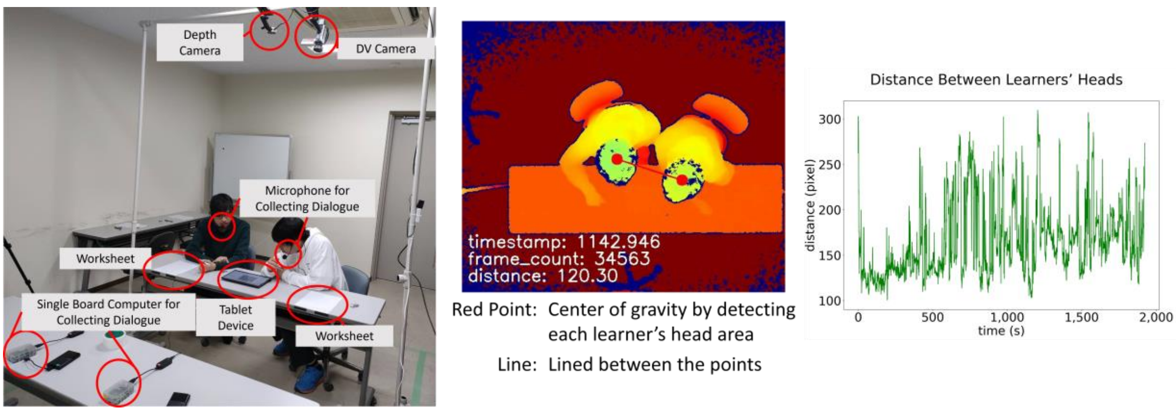



**Figure 1: Outline of the experiment: (left) experimental environment; (middle) depth image of experimental learners captured from above using a depth camera; and (right) a graph of the distance between the learners' heads**

## 3 PRELIMINARY EXPERIMENT

Figure 1 shows an outline of the experiment conducted in this study. Here, the image shows the experimental environment. The proposed method assumes that two learners are sitting side by side against a desk with a tablet device between the learners. In this experiment, we used the tablet device with an interactive math material, and we measured the distance between the learners' heads using a depth camera mounted directly above the learners. The depth camera enables us to perform measurements that are less influenced by the learner's clothing or hair color because the depth camera is not dependent on only color information. We also used digital video cameras to record the experiment, and we attached a microphone to the learners to record their verbal communication. Figure 1 also shows an image of the learners captured by the depth camera from above. Note that this image is annotated by points and a line. The points are drawn to the center of gravity by detecting each learner's head area, and both points are lined. The length of the line indicates the distance between the learners' heads. Figure 1 also shows a graph representing the change in the distance between the learners' heads during the experimental process. Here, the vertical axis represents the distance in pixels, and the horizontal axis represents time (in seconds) from the beginning of learning.

The graph of Figure 1 shows that the proposed method can measure the distance between the learners' heads both automatically and continuously. We also found that the distance between the learners' heads varied during the learning activities, e.g., when the learners are looking or operating the tablet device together, the distance between learners' heads is typically 150 pixels or less, and when writing at the worksheet, the distance between learners' heads is typically 200 pixels or more. In the future, we intend to evaluate the graph further by comparing it to an expert's evaluation results and the learners' internal states. We also intend to combine the distance between the learners' heads and other multimodal information represented by dialogue or joint gazing to identify further improvements to the method.

## ACKNOWLEDGEMENTS

This work has been partly supported by the Grants-in-Aid for Scientific Research (Nos. 21K02752 and 22K02951) by MEXT (Ministry of Education, Culture, Sports, Science and Technology) in Japan.

## REFERENCES

Spikol, D., Ruffaldi, E., Dabisias, G., & Cukurova, M. (2018). Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. *Journal of Computer Assisted Learning*, 34(4), 366–377. https://doi.org/10.1111/jcal.12263

Cukurova, M., Zhou, Q., Spikol, D., & Landolfi, L. (2020). Modelling collaborative problem-solving competence with transparent learning analytics: Is video data enough? *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 270–275. https://doi.org/10.1145/3375462.3375484

Funabashi, R., Nabetani, K., Noda, T., Kaneko, M., & Egi, H. (2022). Motivation estimation method for computer supported collaborative learning using tablet computer, *APSCE the 30th International Conference on Computers in Education*, 2, 671–673. https://icce2022.apsce.net/uploads/P2_WIPP_101.pdf

# Learn from Your First-Year Peers? The Impact of Joint Course Enrollments on Student Performance and Major Choice

**Anastasiia Alekseeva, Celine Onyemaechi, Christian Fischer**
University of Tübingen, Germany
christian.fischer@uni-tuebingen.de

**ABSTRACT**: This study investigates first-year co-enrollment networks and their effect on the final college GPA and major switching. This study uses data from one cohort of students in a large public research university in the U.S. (N = 4,661) and applies a novel method based on student co-enrollment network features based on the nearest 2, 4, 8, and 16 neighbors. We found that accounting for the co-enrollment structure improves model fit, reaching a ceiling at around eight neighbors. Utilizing regression models, we found that students' final GPA was not associated with the demographic characteristics of first-year peers. Citizenship status, performance and major of first-year peers are associated with major changes.

**Keywords**: higher education, GPA, major change, co-enrollment network, nearest neighbors

## 1 INTRODUCTION

Students' social networks are often seen as a key element for student success. One of the ways to understand the impact of the network structure is to examine students' joint enrollment (Gardner et al., 2018). This often induces informal student-organized learning networks built for information exchange, especially in the first year of college (Brouwer et al., 2022). While the effect of informal networks on academic performance is relatively well-studied, the co-enrollment networks themselves, despite being rich in information, and their impact on major change are generally overlooked. Notably, student demographics such as gender and ethnicity or program choices (e.g., STEM major) were shown to be associated with academic performance and major change (Astorne-Figari & Speer, 2019; Tomás-Miquel et al., 2016). The results of this study may support the development of course selection tools and student/instructor support systems. It aims to answer the following research questions (RQs): **(RQ1)** How are first year student enrollment characteristics, as manifested through joint course enrollment, associated with students' final college GPA? **(RQ2)** How are first year student enrollment characteristics, as manifested through joint course enrollment, associated with students' decisions to change their academic major?

## 2 METHODS

### 2.1 Study Settings

This study is situated at a large public research university in California that is federally designated as a Hispanic-Serving Institution (HSI) and an Asian American and Native American Pacific Islander-Serving Institution (AANAPISI) and enrolls a diverse body of about 25,000 undergraduate students. Data was provided by campus offices including Admissions, the Registrar's Office, and the Office of Institutional Research, among others. This study uses data from one cohort of degree-seeking freshmen and sophomore students, who were admitted in the fall 2016 term and graduated within six years of their college admission (N = 4661 students). Missing data was assumed missing completely at random and treated with list-wise deletion.

### 2.2 Measures

Dependent variables in the study are final college GPA (continuous) and an indicator of the occurrence of a major change (dichotomous). Independent variables include dichotomous predictors such as gender, belonging to an underrepresented minority, low-income background, U.S. citizenship, and STEM major indicator, as well as such continuous variables as students' first year and high school GPAs and a scaled admission score variable. In addition, we generated network variables for each predictor by aggregating the attributes of the neighboring nodes in the first-year undirected co-enrollment network. We defined the neighborship in terms of the number of courses students passed together. We selected top 2, 4, 8 and 16 such neighbors and computed averages of their first year and high school GPAs, admission scores, and the percentage of students with the same major and background.

### 2.3 Analytical Methods

We estimated ten model specifications by utilizing OLS regression (RQ1) and logistic regression (RQ2) models and compared the respective models to baselines via F-tests and likelihood ratio tests. The baseline models do not contain the aggregate network characteristics. To control the group-level heterogeneity we introduced the fixed effect on the school of the first major. Moreover, we used robust standard errors given potential clustering effects. As a robustness check, we estimated a set of additional model specifications without some of the predictors, which resulted in similar results.

## 3 RESULTS

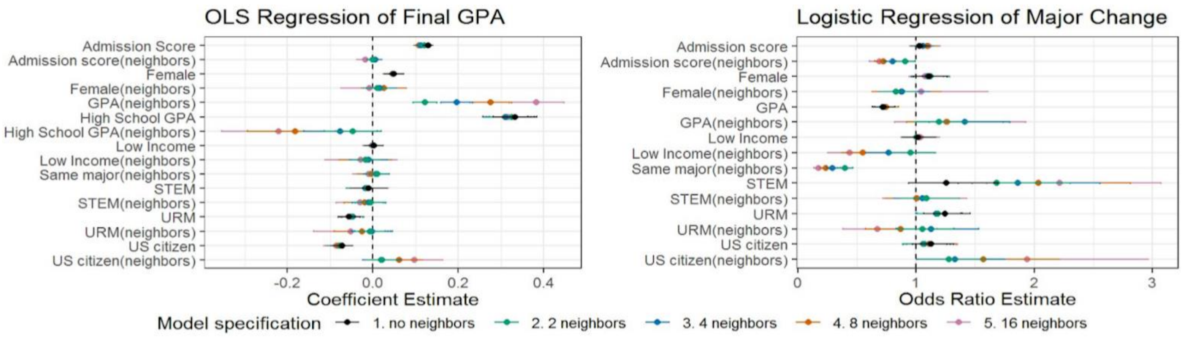
Table 1 shows that information from students' coursemates through joint enrollments significantly improves the prediction of the final GPA and major change. Additionally, we found that adding information about a larger number of neighbors improves the model fit but the relative enhancement of the performance tends to converge at 8 neighbors.

**Table 1. Model performance metrics and results of statistical tests.**

| | **no neighbors** | **2 neighbors** | **4 neighbors** | **8 neighbors** | **16 neighbors** |
|---|---|---|---|---|---|
| **OLS Regressions of Final GPA** | | | | | |
| R Squared | 0.242 | 0.256 | 0.266 | 0.272 | 0.273 |
| p-value (F-test) | | $<10^{-14}$ | $<10^{-26}$ | $<10^{-34}$ | $<10^{-37}$ |
| **Logistic Regressions of Major Change** | | | | | |
| Log-likelihood | -3070.167 | -2999.960 | -2981.776 | -2969.823 | -2956.431 |
| p-value (LRT) | | $<10^{-25}$ | $<10^{-33}$ | $<10^{-38}$ | $<10^{-43}$ |

Examining the linear and logistic regression results indicates that some of the network variables have non-zero associations with the outcome variables (Figure 1). RQ1: Aggregate GPA of neighbors have a significant positive association (0.12 - 0.38) with students' final GPA, which could be explained by heterogeneity of grades in different courses; the significance of citizenship status and final GPA differs with the number of neighbors. RQ2: Higher GPA of a student in the first year decreases the odds ratio by 0.72 - 0.74, while admission score of the neighbors decreases it by 0.68 - 0.90. Also, if most of the neighbors have the same major as the student, students will be less likely to change their major, while citizenship status of the neighbors shows an opposite effect.

**Figure 1. Coefficient estimates with 95% confidence intervals. Intercept and fixed effects are omitted from both graphs.**




# 4 DISCUSSION

This study contributes to research on co-enrollment networks and introduces a new approach to incorporating the network information. The administrative data used in this study is typically available at every university (Fischer et al., 2020), which allows institutional researchers to replicate our analyses in their individual contexts. The analysis code used in this study is provided in our public GitHub repository: https://github.com/ana-alekseeva/lak_network_paper. Overall, the first-year co-enrollment network features help explain undergraduate students' performance (RQ1) and major change (RQ2). Furthermore, it allows finding an "effective" number of neighbors to consider in statistical models. Further research could increase the scope of the dataset by adding more cohorts from the same (and other) universities, compare network structures of transfer and traditional students, and apply more complex models such as convolutional graph neural networks.

## REFERENCES

Astorne-Figari, C., & Speer, J. D. (2019). Are changes of major changes? The roles of grades, gender, and preferences in college major switching. *Economics of Education Review*, *70*, 75-93.

Brouwer, J., De Matos Fernandes, C. A., Steglich, C. E. G., Jansen, E. P. W. A., Hofman, W. H. A., & Flache, A. (2022). The development of peer networks and academic performance in learning communities in higher education. *Learning and Instruction*, *80*, 101603.

Fischer, C., Pardos, Z. A., Baker, R. S., Williams, J. J., Smyth, P., Yu, R., Slater, S., Baker, R., & Warschauer, M. (2020). Mining Big Data in Education: Affordances and Challenges. *Review of Research in Education*, *44*(1), 130–160.

Gardner, J. P., et al. (2018). Learn From Your (Markov) Neighbor: Coenrollment, Assortativity, and Grade Prediction in Undergraduate Courses. *Journal of Learning Analytics*, *5*(3), 42–59.

Tomás-Miquel, J.-V., Expósito-Langa, M., & Nicolau-Juliá, D. (2016). The influence of relationship networks on academic performance in higher education: A comparative study between students of a creative and a non-creative discipline. *Higher Education*, *71*(3), 307–322.

# Further Evidence for Regularity in Student Learning Rates Across Demographic, Academic Proficiency, and Motivational Groups

**Gillian Gold**
Yale University
gillian.gold@yale.edu

**Conrad Borchers**
Carnegie Mellon University
cborcher@cs.cmu.edu

**Paulo Carvalho**
Carnegie Mellon University
pcarvalh@andrew.cmu.edu

**ABSTRACT**: To replicate and expand on previous results showing that student learning rates are regular under favorable learning conditions (Koedinger et al., 2023), we used a dataset of 426 students who engaged with a cognitive tutoring system throughout an academic year. We used the individual additive factors model (iAFM) to estimate student parameters: intercept (initial knowledge) and slope (learning rate). Our findings replicate regularity in learning rates, including across student subgroups defined by sex assigned at birth, socioeconomic status, academic proficiency, and self-reported measures of motivation. Moreover, initial knowledge within subgroups was positively correlated with academic performance and self-reported goal orientation at the onset of the school year. There were no significant correlations found between learning rate and demographic or motivational measures. One important implication of these findings is that interventions should target prior knowledge and availability of additional practice opportunities for struggling students, along with motivational support so that students seek those external additional opportunities.

**Keywords**: tutoring systems, student motivation, learning rate, cognitive modeling, K-12

## 1 INTRODUCTION

Intelligent tutoring systems (ITS), such as cognitive tutors, create favorable learning conditions by providing deliberate practice, feedback, and step-by-step instruction (Koedinger et al., 2023). By recording fine-grained student interactions with learning systems, ITS data also allows for modeling student learning and knowledge growth (Koedinger et al., 2023). Koedinger et al. (2023) found a lack of variation in learning rates, measured by increases in correctness after each practice opportunity of given Knowledge Components (KCs). However, there may be learning variation within specific student subgroups. Past research has demonstrated that diligent students tend to engage more readily with learning tasks, while less diligent students often fail to fully utilize available learning opportunities (Bernacki et al., 2013). Taken together, we hypothesized that, when exploring different subgroups, academic proficiency and goal orientation would be related to increased variation only in students' initial knowledge. If the cognitive tutor creates optimal learning conditions (Koedinger et al., 2023), it will help all students learn at the same rate, but differences in initial knowledge will contribute to differences in academic proficiency and motivational approaches.

## 2 METHODOLOGY

### 2.1 Data Collection

We retrieved the dataset (ds613) used in this analysis from DataShop (Koedinger et al., 2010). The data was collected over a year-long study in a suburban middle school in the mid-Atlantic region of the United States. Students utilized the Carnegie Learning Cognitive Tutor software (CogTutor). The dataset comprises math practice learning transactions of 426 6-12th grade students who used the tutor for approximately two class periods per week for 8 months. It also includes demographic information, academic proficiency measures, and motivational survey responses (Bernacki et al., 2013). 155 students had no accompanying demographic or external achievement information, so 271 students were included in the analysis.

## 2.2 Analysis Methods

We used the individual additive factors model (iAFM) to calculate individual student parameters: intercept represents initial knowledge and slope represents learning rate (Koedinger et al., 2023). Student subpopulations were based on demographics (sex assigned at birth, socioeconomic status), academic proficiency (previous final grade, final math grade, state math exam), and self-reported motivational measures (mastery approach, performance approach, performance avoidance) (Bernacki et al., 2013). We followed the same approach as in Koedinger et al. (2023) to calculate the number of opportunities to reach mastery. To test whether variability in learning rate and intercept within subgroups, we divided students into high and low groups (using median split) for each measure of interest, and then compared opportunities to reach 90% mastery across students with low/high initial knowledge and those with low/high learning rate. A larger difference suggests higher variability (see Koedinger et al., 2023). To test for relationships between each measure of interest and initial knowledge or learning rate, we used Pearson's correlation. We hypothesized that learning rate would not vary substantially even within sub-groups and that proficiency and motivational measures would correlate more with initial knowledge than learning rate.

# 3 RESULTS

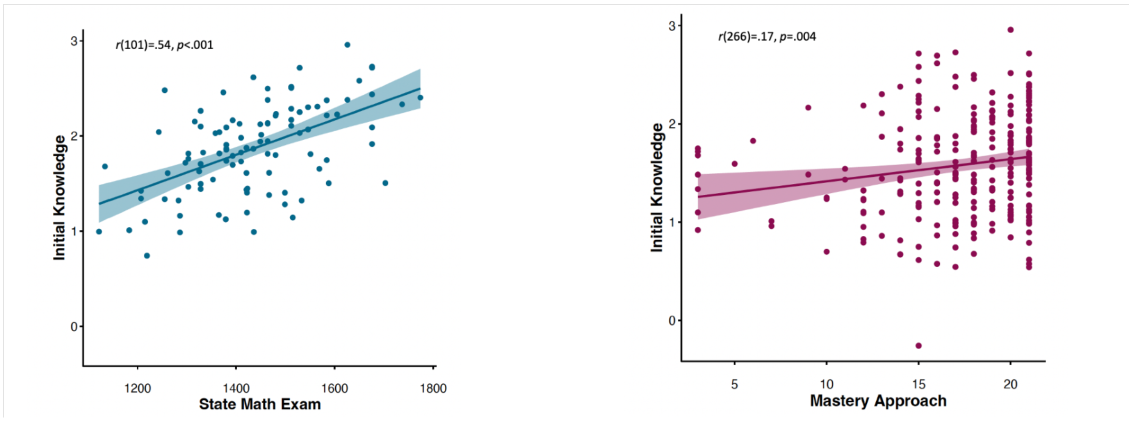
We found that, across subgroups in the academic and motivational measures, initial knowledge varied considerably. For example, students who rated high on mastery approach and had high initial knowledge took on average 4.41 opportunities to reach mastery, whereas those who had lower initial knowledge took on average 12.20 opportunities: a difference of 7.79 opportunities. However, for learning rate, we found small variation in reaching mastery (0.37 opportunities). Table 1 presents the difference in learning opportunities to reach mastery between high and low initial knowledge/learning rate students for selected subgroups of students.

**Table 1: Difference in number of opportunities to reach mastery between high and low initial knowledge/learning rate for different sub-groups of students.**

| | State Math Exam | | Mastery Approach | |
|---|---|---|---|---|
| | Low | High | Low | High |
| **(High-Low) Initial Knowledge** | 6.30 | 5.40 | 8.33 | 7.79 |
| **(High-Low) Learning Rate** | 0.33 | 0.42 | 0.39 | 0.37 |

There were no statistically significant differences in initial knowledge or learning rate for students based on sex assigned at birth ($r(267)=.08$, $p=.170$) or eligibility for free/reduced lunch ($r(268)= .04$, $p=.490$). Higher initial knowledge was significantly associated with higher prior final grade ($r(238)=.40$, $p<.001$), final math grade ($r(254)=.44$, $p<.001$), and state math exam ($r(101)=.54$, $p<.001$) (Figure 1). Moreover, higher initial knowledge was significantly associated with higher reports of mastery approach ($r(266)=.17$, $p=.004$) (Figure 1), performance approach ($r(266)=.13$, $p=.038$), and performance avoidance (*r*(266)=.20, *p*<.001). There were no significant differences in learning rate in subgroups based on academic proficiency or achievement goals (all *r*'s<.12, *p*'s>.05).

**Figure 1: Left: Positive correlation between initial knowledge and final state math exam scores (95% CI [0.39, 0.66]). Right: Positive correlation between initial knowledge and self-reported mastery approach at the beginning of the school year (95% CI [0.06, 0.29]).**




## 4 DISCUSSION

This study supports the hypothesis that learning rates are generally regular across students (Koedinger et al., 2023). Consistent with Koedinger et al.'s findings and theoretical interpretation, prior knowledge varied substantially: a student with low prior knowledge requires about 8 more opportunities to reach mastery than a student with high prior knowledge, regardless of their mastery approach. Comparatively, learning rate varied very little. However, and importantly, we found positive correlations between initial knowledge and both academic proficiency and motivational measures. These findings suggest that prior knowledge may contribute to higher variability in learning outcomes. If a student is highly motivated, they might have sought more opportunities in the past, begin with a higher level of knowledge, and perform better at the start of the year. However, higher motivation does not seem to affect the speed at which students learn. Provided that the learning conditions are favorable, as in the cognitive tutor used in this study, all students will learn at approximately the same rate. Thus, potential interventions should target initial knowledge by providing additional practice opportunities, which may also require motivational support for struggling students.

## REFERENCES

Bernacki, M. L., Nokes-Malach, T. J. & Aleven, V. (2013). Fine-grained assessment of motivation over long periods of learning with an intelligent tutoring system: methodology, advantages, and preliminary results. In R. Azevedo, & V. Aleven (Eds.) *International Handbook of Metacognition and Learning Technologies*.

Koedinger, K.R., Baker, R.S.J.d., Cunningham, K., Skogsholm, A., Leber, B., Stamper, J. (2010) A Data Repository for the EDM community: The PSLC DataShop. In Romero, C., Ventura, S., Pechenizkiy, M., Baker, R.S.J.d. (Eds.) Handbook of Educational Data Mining. Boca Raton, FL: CRC Press.

Koedinger, K. R., Carvalho, P. F., Liu, R., & McLaughlin, E. A. (2023). An astonishing regularity in student learning rate. Proceedings of the National Academy of Sciences of the United States of America, 120(13), e2221311120. https://doi.org/10.1073/pnas.2221311120

# Understanding Learners' Cross-context Self-direction Skill Achievement Behavior

**Junya Atake[1], Chia-Yu Hsu[1], Huiyong Li[2], Izumi Horikoshi[1], Rwitajit Majumdar[3], Hiroaki Ogata[1]**
[1]Kyoto University, [2]Kyushu University, [3]Kumamoto University
atake.junya.86t@st.kyoto-u.ac.jp

**ABSTRACT**: This study examined whether self-direction skills (SDS) are generic and can be applied to a variety of situations using data, rather than being limited to a specific context. The concept of SDS, a crucial component in 21st-century learning, encompasses activities ranging from academic learning to daily physical tasks. The GOAL system was developed to overcome the challenge of collecting and synchronizing data from learners' daily activities. The following research question was answered: How do learners' self-direction skills achievement behaviors differ in the contexts of physical activities and learning activities? To answer this question, log data from the GOAL system of 120 Japanese junior high school students, from June 2020 to March 2023, were analyzed. Sankey diagrams visualized the highest SDS scores achievement patterns of learners across various physical and learning contexts. The findings revealed that learners who scored well in the initial contexts tended to maintain similar achievements in subsequent contexts. This study supports previous research by providing new evidence on skill transfer from trace data analyses.

**Keywords**: Learning Analytics, Self-Direction Skills, GOAL system, Cross-context

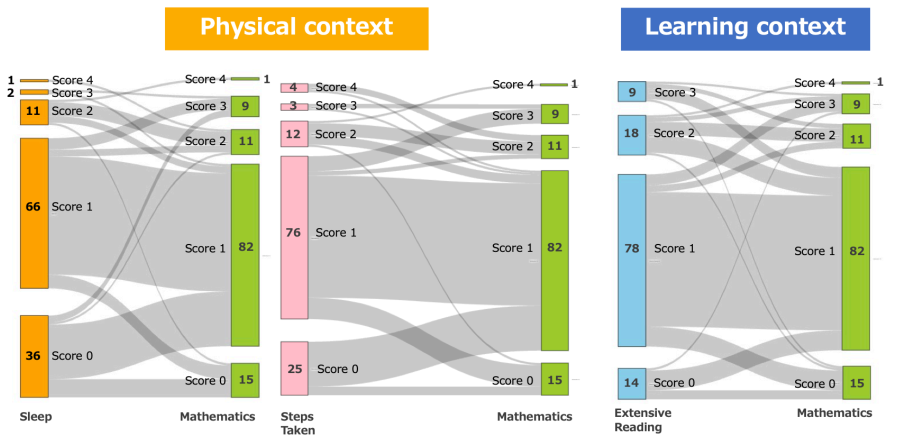
## 1 INTRODUCTION

Self-direction Skill (SDS) is a meta-skill in the Self-directed Learning (SDL) process, where learners have the autonomy to diagnose their learning needs, set goals, select strategies, and evaluate outcomes. Learners' acquisition of SDS may occur not only in learning but also in physical activities such as daily exercise and sleeping (Yang et al., 2023). Previous studies have argued that skills are generic and transferable across content areas (Budge, 2000; Brandt, 2020). The GOAL system was developed based on this premise (Majumdar et al., 2018). This system was developed to understand the complex and different contexts of learning activities and exercises, to make SDL effective and support planning, execution, and reflection using the DAPER process model (Majumdar et al., 2018). This study aims to empirically determine whether SDS acquisition behavior is generic and can be applied to a variety of contexts, rather than being limited to a specific context. With the available data from the GOAL system, we answer the following research question: How do learners' self-direction skills achievement behaviors differ in the contexts of physical activities and learning activities? Answering this question contributes to deepen our understanding of the transferability of SDS, along with discussions regarding the technological support for acquiring SDS in new contexts.

## 2 LEARNING CONTEXT AND DATA ANALYSIS

We used SDS scores of the GOAL system, implemented in Japanese junior high schools from June 2020 to March 2023. During this period, 39765 logs were collected from 120 learners. In the GOAL system, SDS were observed in various contexts. Specifically, the system calculates SDS scores in contexts such as Extensive Reading, Mathematics, Steps Taken, and Sleep. Extensive Reading and Mathematics contexts are partly teacher-assisted activities; learners are supported in their activities by teachers (e.g., the teacher introduces the activity and instructs how to conduct it). In contrast, Steps and Sleep are completely learner-centered activities; the teacher only distributes wearable devices to learners who are responsible for deciding what they will do. The SDS score was calculated for each of the five DAPER phases: collecting, analyzing, planning, monitoring, and reflecting. In this study, we only visualized SDS during the analysis phase because it is a vital phase in determining whether learners can plan properly (Yang et al., 2021). The skill level for the analysis phase was calculated from 0 to 4 based on the number of actions and the accuracy of the analysis report. Learners analyze activity trends by checking visualized self-data and related data (e.g., the average/maximum/minimum of the group data) and evaluate their current activity status (e.g., either good or bad) (Yang et al., 2023). A score of 0 revealed that the learner did not perform any analysis. A score of 1 suggested that the learner checked data but did not identify the status. A score of 2 demonstrated that the learner checked the data but did not identify the situations other than the information provided by the system. A score of 3 indicated that the learner checked the data and successfully identified the information provided by the system. Finally, a score of 4 displayed that the learner checked the data without relying on the information provided by the system and successfully identified their status. We regarded the highest SDS as a learner's achievement within a specific period. Sankey diagrams were used to visualize transitions between scores across four contexts—Sleep, Steps Taken, Extensive Reading, and Mathematics. Each node represents the number of people with each score. In this dataset, the Mathematics context was executed in the latter half of the period. Therefore, the nodes were set to the right as a later activity. On the other hand, Sleep, Steps Taken, and Extensive Reading were placed on the left as initial activities. Each flow shows how the score in one context is transferred to another context.

## 3 RESULT AND DISCUSSION




| (a) Total | n=116 | n=118 | n=118 |
|---|---|---|---|
| (b) Remain* | 68 (59%) | 69 (58%) | 72 (61%) |
| (c) Improve* | 36 (31%) | 36 (26%) | 21 (18%) |
| (d) Downgrade* | 12 (10%) | 13 (16%) | 25 (21%) |

*The Number and Ratio(%) of learners who remain in the same score level (b) , who improve (c) and downgrade(d).

**Figure 1: Sankey diagram of skill transfer in four contexts**

The Sankey diagram (Figure 1) illustrates the transition of learners' SDS acquisition scores across contexts in the GOAL system. This indicates that scores obtained in the contexts to which the learners were exposed initially (Steps, Sleep, and Extensive Reading) influenced their scores in later contexts (Mathematics). According to the Sankey diagram (Figure 1), learners who scored 1 or 2 in each initial context (Steps Taken, Sleep, and Extensive Reading) also achieved a score of 1 or 2 in the later context (Mathematics) (over 80%). On the contrary, most of the learners with scores of 3 (n=9) and 4 (n=1) in the initial context maintained high scores in the later context. Some learners with scores of 1 and 2 achieved higher scores in the later context (overall 10-20%). These findings suggest that learners tend to achieve SDS, showing similar patterns in both physical and learning activities within the GOAL system. In addition, we can observe a tendency for initial skill levels to improve or downgrade in later contexts. The potential factors that indicate different trends could be the levels of learners' intrinsic motivation and teachers' support for the particular activity. These potential factors need to be further investigated.

## 4 CONCLUSION

In this study, we examined whether the SDS is generic and can be applied to various contexts supported by the GOAL system. The results suggested that learners' SDS were generic; those acquired in a specific context of physical and learning activities were transferable within the GOAL system. This supports the claims of previous studies (Brandt, 2020; Budge, 2000) and provides new evidence from trace data. However, this study only examined the SDS during the analysis phase. In addition, we did not compare teacher-assisted and learner-centered activities. Future research should consider these aspects in greater detail and specifically extend opportunities for SDL training in daily activities with GOAL including academic pursuits (e.g., science and coding activities) and school club activities (e.g., sports and music).

## 5 ACKNOWLEDGEMENT

This work is partially funded by JSPS KAKENHI JP22H03902, JSPS KAKENHI JP23H00505.

## REFERENCES

Brandt, W. C. (2020). *Measuring student success skills: A review of the literature on self-directed learning*. National Center for the Improvement of Educational Assessment.

Budge, D. (2000). Motivating students for lifelong learning. Education and Skills. (ERIC No. ED449566)

Majumdar, R., Yang, Y., Li, H., Akçapınar, G., Flanagan, B., & Ogata, H. (2018). GOAL: A system to support learner's acquisition of self-direction skills. *In Proceedings of the 26th International Conference on Computers in Education* (pp. 406-415).

Yang, Y. (2023). Facilitating contextual self-directed learning by using GOAL system in K-12 education. [Unpublished doctoral dissertation]. Kyoto University.

Toh, W., & Kirschner, D. (2020). Self-directed learning in video games, affordances and pedagogical implications for teaching and learning. *Computers and Education*, 103912.

Yang, Y., Majumdar, R., Li, H., Akçapinar, G., Flanagan, B., & Ogata, H. (2021). A framework to foster analysis skill for self-directed activities in data-rich environment. Research and Practice in Technology Enhanced Learning, 16, 1-17.

# Effects of Learning Analytics Support on Self-Regulated Learning Strategies in Virtual Reality Content Creation

**Jeremy Tzi Dong Ng[1], Xiao Hu[1], Yizhou Fan[2], Dragan Gašević[3]**
[1]The University of Hong Kong, [2]Peking University, [3]Monash University
jntd@connect.hku.hk, xiaoxhu@hku.hk, fyz@pku.edu.cn, dragan.gasevic@monash.edu

**ABSTRACT**: Learning analytics (LA) holds the promising potential of facilitating learners' self-regulated learning (SRL) during maker activities, though the impacts of LA support on learners' SRL strategies in maker activities remain under-explored. Thus, this study conducted a classroom quasi-experiment with 101 students in a university-wide general education course, with LA support in the form of a student-facing dashboard as the intervention. Both logs from a VR creation platform and interview data were collected. Based on a maker-based SRL analytic framework, we compared frequencies of macro-level and micro-level SRL actions between the experimental (i.e., with LA support, N=52) and control groups (N=49). Supplemented with interview responses, results revealed that, 1) the availability and usage of LA support correlated with several SRL actions, and 2) the availability of the LA support benefited students' self-reflection. Implications were drawn for research and practices in maker education and learning analytics.

**Keywords**: Maker education, Self-regulated learning, learning analytics dashboard, Virtual reality content creation

## 1. INTRODUCTION

Premised on empowering student agency, maker activities involve learners in creating public-facing products (e.g., digital stories) for acquiring content knowledge and skills (Ng et al., 2022; Ng et al., 2023). The iterative procedures of ideation, experimentation, and reflection on ideas, tasks, and actions also characterize maker activities (Ng et al., 2023), which align well with common self-regulated learning (SRL) strategies (e.g., formulation, reflection) (Zimmerman, 2002). LA dashboards have the potential in generating actionable feedback for promoting SRL (Yilmaz & Yilmaz, 2020), and help learners navigate through the exploratory trial-and-error process of maker activities (Ng et al., 2023). Nonetheless, few studies have investigated whether and how LA support can facilitate learners' SRL strategies during maker activities. To fill the gap, this study conducted a classroom quasi-experiment to examine the effects of LA support on students' SRL strategies during the emerging maker activity of virtual reality (VR) content creation. We posed the research question: *What are the effects of LA support on students' SRL strategies in VR content creation?*

## 2. RESEARCH CONTEXT AND METHODS

A classroom quasi-experiment was conducted in an online undergraduate general education course on digitizing cultural heritage in Spring 2022 at a comprehensive university in Hong Kong. As a major assignment, a maker activity was delivered in which students created an individual VR story for showcasing a cultural heritage (e.g., a temple). Designed based on the maker-based pedagogy, the maker activity involved three consecutive week-long stages, including 1) authoring draft VR content,

2) peer evaluation of drafts, and 3) preparing a final, public-facing version. A tutor led hands-on practices across three weekly tutorials in 10 tutorial classes. A total of 101 students (62% female, 49% first-year), from both STEM (50%) and non-STEM majors (50%), gave informed ethical consents.

The maker activity of VR creation was conducted on our self-developed web-based platform, LAVR (Figure 1a), where student-creators can add or edit VR scenes, multimedia objects (e.g., text, images, audio) and their properties (e.g., volume), as well as view and review peers' creation. Based on student needs (Ng et al., 2022), we also developed a student-facing LA dashboard on LAVR (Figure 1b). Intended to promote students' SRL, this dashboard is composed of a checklist of tasks of the maker activity (e.g., "add text object") and a page displaying aggregated statistics of components in the VR stories (e.g., average number of text objects). In our quasi-experiment, the ten tutorial classes (n = 101) were randomly assigned to the control (five classes, 49 students in total) (49%) and experimental (other five classes of 52 students) (51%) groups. While all students created VR stories using LAVR, for the intervention on LA support, only the experimental group could access the dashboard, and was encouraged but not required to use it. The teaching team did not have access to any information shown on the dashboard.

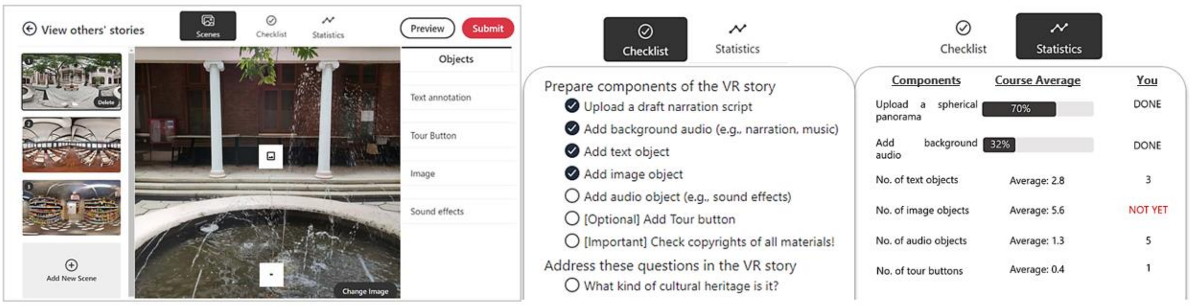



**Figure 1: (a) Editor interface (left) and (b) Learning analytics dashboard (right) on LAVR**

This study collected system logs as the main data source (N = 101), supplemented by semi-structured interview data (N = 32). We extracted a total of 69,612 log events from LAVR between February and early March 2022. From these raw logs, we identified 56 kinds of micro-level actions (e.g., Edit Text Object) as meaningful learning actions, which were categorized into each of 11 macro-level actions (e.g., Edit Component), that could be more generalizable to other kinds of maker activities and more explainable to teachers and learners. With reference to current log-based SRL research (e.g., Fan et al., 2021), an analytic framework (Table 1) was developed, adapting common SRL strategies as conceptualized in the literature. Then, to operationalize these strategies, we mapped each of the macro-level actions (e.g., Enter Platform) to one SRL strategy (e.g., Goal-setting). To ensure the validity of our interpretations, we carefully designed and iteratively refined the analytic framework.

**Table 1: Analytic framework of maker-based SRL strategies, macro-level and micro-level actions**

| SRL strategy | Definition | Macro-level action | Micro-level action |
|---|---|---|---|
| Goal-setting | To define goals of the maker activity. | Enter Platform | Enter Assignment |
| Observation | To observe one's draft or product. | Open/Close Panel | Enter Review List |
| Formulation | To engage in creating new components. | Add Component | Add Scene |
| Reformulation | To edit existing components. | Edit Component | Edit Text Object |
| Reflection | To reflect on progress and performance. | Write Peer Reviews | |

## 3. RESULTS

Results of non-parametric Mann-Whitney U tests revealed that the frequencies of macro-level SRL actions in the experimental and control groups exhibited no significant differences ($0.062 < p < 0.686$). On the micro-level, students in the experimental group performed the actions “Enter Assignment” ($p = 0.023$*) (corresponding to the SRL strategy, Goal-setting) and “Enter Review List” ($p = 0.041$*) (Observation) less frequently than those in the control group, but were more frequent in “Open Overview Panel” ($p = 0.044$*) (Observation). In the experimental group, the more frequently students opened their checklist (Figure 1b) for “easily missed details” (#31), the more frequently they would enter the full-screen mode of their VR stories ($r = 0.279$; $p = 0.045$*) (i.e., Observation). Results also showed positive correlations between students’ frequency of checking items on the checklist and their frequencies of deleting a text object ($r = 0.306$; $p = 0.027$*), uploading a background audio ($r = 0.331$, $p = 0.016$*), changing the background audio’s volume ($r = 0.303$; $p = 0.029$*), and deleting a background audio ($r = 0.347$; $p = 0.012$*) in the VR story, respectively (i.e., Reformulation). Students in both groups were inclined to “focus on the quality of their own [creation]” (#1) and “not interested in others’ progress” (#8). Nevertheless, those in the experimental group perceived that the Statistics page of the LA dashboard (Figure 1b) were helpful in “reminding [them] what [aspects] to improve” (#15) (Reflection), implying the potential of LA-enabled feedback in helping students embrace collective, peer-generated innovation arising from maker activities (Ng et al., 2022).

## 4. CONCLUSION AND FUTURE WORK

This study revealed the effects of LA support on micro-level SRL action frequencies in the maker activity of VR creation. Methodologically, this study leveraged raw logs as a quantitative data source for evaluating making processes, which were often qualitatively assessed. Theoretically rooted in SRL, the macro-level SRL actions derived from our analytic framework can be applied across other relevant learning contexts for analyzing the underlying self-regulatory mechanisms. For practical implications, this study demonstrates the feasibility of introducing LA support to maker activities. Future work will generate the sequences of SRL actions for more nuanced insights into students’ learning processes.

## 5. REFERENCES

Fan, Y., Saint, J., Singh, S., Jovanovic, J., & Gašević, D. (2021). A learning analytic approach to unveiling self-regulatory processes in learning tactics. In *LAK21: 11th International Learning Analytics and Knowledge Conference*, 184-195.

Ng, J. T. D., Liu, Y., Chui, D. S., Man, J. C., & Hu, X. (2023). Leveraging LMS Logs to Analyze Self-Regulated Learning Behaviors in a Maker-based Course. In *LAK23: 13th International Learning Analytics and Knowledge Conference*, 670-676.

Ng, J. T. D., Wang, Z., & Hu, X. (2022). Needs analysis and prototype evaluation of student-facing LA dashboard for virtual reality content creation. In *LAK22: 12th International Learning Analytics and Knowledge Conference*, 444-450.

Yilmaz, F. G. K., & Yilmaz, R. (2020). Student opinions about personalized recommendation and feedback based on learning analytics. *Technology, knowledge and learning, 25*, 753-768.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into practice, 41*(2), 64-70.

# The Impact of Different Personalisation Algorithms on Literacy and Numeracy in Kenyan Pre-primary Education: A Comparative Study of Summative and Formative Assessments Results

**Chen Sun[1], Louis Major[1], Nariman Moustafa[2], Rebecca Daltry[3], Lazar Obradovic,[4] Aidan Friedberg[4]**

[1]University of Manchester, chen.sun@manchester.ac.uk; louis.major@manchester.ac.uk

[2]Open Development & Education, nariman@edtechhub.org

[3]Jigsaw, rebecca@edtechhub.org

[4]EIDU, lazar.obradovic@eidu.com; aidan.friedberg@eidu.com

**ABSTRACT**: Digital personalisation has demonstrated potential to enhance learning. However, there is limited evidence on the comparative impact of different content personalisation algorithms on early-year numeracy and literacy outcomes, especially in low- and middle-income countries (LMIC). This paper reports an A/B/C test conducted over three weeks via a digital personalised learning tool used by 6479 Kenyan pre-primary learners. Two personalisation algorithms were implemented (maximising learner engagement or score), while expert-curated sequence was used as a control. Learners from 1509 classes were randomly divided across the three partitions. Learning in numeracy and literacy was compared across three metrics: summative assessment, curriculum progress, and formative assessment. Results showed no difference between partitions in the summative assessment. Different partitions tended to progress through the digital curriculum at a different pace. Significant differences between partitions were found in formative assessment scores, with the impact of each algorithmic approach varying according to different learning strands. Findings contribute to a deeper understanding of how different algorithms impact pre-primary education in LMIC contexts, with implications for designing personalised learning approaches tailored to specific learning content and learner profiles.

**Keywords**: Digital personalised learning, pre-primary, low- and middle-income country, literacy, numeracy

## 1 INTRODUCTION

Evidence indicates that digital personalised learning (DPL) can have a positive impact on learning outcomes. An important component of personalisation is to sequence learning content to actively engage learners (Diwan et al., 2023) and / or increase knowledge acquisition (Major et al., 2021). Research suggests that content sequencing powered by personalisation algorithms can outperform expert-suggested sequencing (Chau et al., 2018). Among various neural network-based algorithms to sequence content, Long Short Term Memory (LSTM) is commonly used (Huo et al., 2020; Ren & Wu, 2023). Research on LSTM-based algorithms reported two specific purposes: maximising learning outcomes and maximising engagement. This paper contributes to research on digital personalisation in that (1) we implemented and compared two personalisation algorithms (optimising engagement vs. score), evaluated against a default content sequence to assess learning effectiveness, and (2) the study was conducted in an under-researched context, i.e., a low- and middle-income country (LMIC; Kenya). The impact of three different content sequencing methods are investigated, by comparing effects on three learning metrics: summative assessment, curriculum progress, and formative assessment. The main research question is: What is the impact of personalisation (Engagement vs Score vs Expert-curated sequence) on learning for Kenyan pre-primary learners?

## 2 METHOD

The EIDU DPL platform runs on low-cost Android devices. Learning units align with the Kenyan curriculum in domains (Numeracy and Literacy) and strands (e.g., Classification). An A/B/C test was conducted over three weeks in July 2023 in 1619 low-cost private pre-primary schools in Nairobi, Kenya. Ethical consent for the research was obtained from both the Kenyan government and relevant organisational bodies involved in the study. Parental consent was substituted by institutional and teacher consent. Pre-primary learners (3-6 years old) were randomised and split equally between three experiment partitions (Engagement vs. Score vs. Expert-curated sequence) with anonymised account IDs serving as a hash function seed. The LSTM-based algorithms predicted engagement and scores for each learning unit based on vectors of student performance history. The pre-test data were selected from learners who used EIDU in June 2023, by matching anonymised IDs in the post-test data. The pre-test sample consisted of 5884 learners from 1177 classes. The post-test sample was 6479 learners from 1509 classes. The increase in post-test sample is attributed to inclusion of new schools and classes joining the experiment in July 2023. Learners were assessed using established summative assessments (EGMA, EGRA or MELQO; Friedberg, 2023), curriculum progress (total usage and number of unique units completed), and formative assessment (scores from learning units).

## 3 RESULTS

**Summative Assessment**. 1661 learners completed the summative assessment in the Engagement partition, 1702 in the Score partition, and 1640 in the expert-curated sequence group. Possible score range was 0 to 1 for each assessment unit. Scores were averaged across all test units and aggregated to overall scores for literacy and numeracy. ANOVA tests did not reveal significant group differences for pre-test on Literacy learning ($F(2,4106) = 2.96$, $p = .052$) and on Numeracy learning ($F(2,1906) = .11$, $p = .89$). Similarly, post-test analysis did not show any group difference in Literacy ($F(2,4361) = .96$, $p = .38$) and in Numeracy ($F(2,2186) = .58$, $p = .56$).

**Curriculum Progress**. No differences were found in total usage between partitions ($F(2,6488) = 2.00$, $p = .13$), although there were significant differences in the number of unique learning units completed ($F(2,6488) = 1509.58$, $p < .001$). See Table 1. The Engagement partition progressed through the highest number of unique units. The Engagement partition solved significantly more unique units (13.45) than the Score partition and significantly more units (3.11) than the expert-curated group. The Score partition, however, completed significantly fewer units (10.34) than the expert-curated partition.

**Table 1: Curriculum Progress: Usage and Number of unique learning units completed.**

| | Total duration (in minutes) | Average progress (*Mean*) |
|---|---|---|
| Engagement | 96.73 | 21.6 |
| Score | 99.73 | 8.2 |
| Expert-curated sequence | 103.74 | 18.5 |

**Formative Assessment**. ANOVA comparing the three groups in the pre-test sample showed no significant differences between the partitions across all eight strands, suggesting the post-test sample is comparable. In total, 6371 learners participated across all three partitions: 2089 Engagement, 2117 Score, and 2165 expert-curated. They collectively played 216 common learning units. ANOVA revealed group differences occurred within all nine learning strands. Analysis using Tukey post hoc tests allowed for pairwise group comparisons (Table 2), suggesting different personalisation strategies may benefit learning in different ways, depending on learning strand or measure in question. The results corroborate the positive effects of content sequencing literature.

**Table 2: Post hoc Tukey test for group comparisons**

| Partition | Eg vs. Expert Mean (SE) | Score vs. Expert Mean (SE) | Eg vs. Score Mean (SE) |
|---|---|---|---|
| Classification | .033 (.004) *** | .044 (.004) *** | -.011 (.003) ** |
| Listening | .028 (.007) *** | -.068 (.007) *** | .096 (.006) *** |
| Measurement | .056 (.008) *** | .042 (.008) *** | .014 (.009) |
| Numbers | -.044 (.008) *** | -.162 (.008) *** | .119 (.009) *** |
| Phonological awareness | -.017 (.007) | -.022 (.007) ** | .006 (.007) |
| Reading | -.013 (.008) | -.117 (.008) *** | -.104 (.007) *** |
| Speaking | .043 (.007) *** | -.034 (.007) *** | .076 (.006) *** |
| Writing | -.001 (.001) | -.005 (.001) *** | .006 (.001) *** |

Note: ** p < .01, *** p < .001. Eg = Engagement, Score = Score partition, Expert = expert-curated

## 4 DISCUSSION AND FUTURE WORK

This work contributes to a deeper understanding of how low-cost DPL benefits literacy and numeracy learning for pre-primary learners in LMICs. The findings demonstrate varied effects of different content sequencing algorithms on specific learning content. Personalisation had no impact on the summative assessment, but may affect learning pathways (e.g., Engagement partition went through learning units faster) and improve certain content learning. Future research should focus on investigating and identifying algorithms that are more beneficial for pre-primary learners in LMICs, taking into account the specific subject matter. Further investigation is needed to pinpoint the exact effects of content sequencing algorithms, by comparing different LSTM-based algorithm designs.

## REFERENCES

Diwan, C., Srinivasa, S., Suri, G., Agarwal, S., & Ram, P. (2023). AI-based learning content generation and learning pathway augmentation to increase learner engagement. *Computers and Education: Artificial Intelligence*, *4*, 100110.

Friedberg, A. (2023). Can A/B Testing at Scale Accelerate Learning Outcomes in Low- and Middle-Income Environments?. In (Eds.) *Communications in Computer and Information Science*: *Vol. 1831* (pp. 780-787). Springer.

Huo, Y., Wong, D. F., Ni, L. M., Chao, L. S., & Zhang, J. (2020). Knowledge modeling via contextualized representations for LSTM-based personalized exercise recommendation. *Information Sciences*, *523*, 266-278.

Major, L., Francis, G. A., & Tsapali, M. (2021). The effectiveness of technology-supported personalised learning in low- and middle-income countries: A meta-analysis. *British Journal of Educational Technology*, 52, 1935–1964.

Ren, J., & Wu, S. (2023). Prediction of user temporal interactions with online course platforms using deep learning algorithms. *Computers and Education: Artificial Intelligence*, *4*, 100133.

# Analysis of learning behavior in an English vocabulary-learning support system with Automatic Speech Recognition

**Saki Hirata[1], Masanori Yamada[2]**
Graduate School of Human-Environment Studies, Kyushu university[1]
Division of Digital Society Research, Data-Driven Innovation Initiative, Kyushu university[2]
rilaxsaki@mark-lab.net, mark@mark-lab.net

**ABSTRACT**: Pronunciation is important In English vocabulary learning. To address this issue, we developed an English vocabulary-learning support system using speech recognition technology. To verify the effect of speech recognition technology on English vocabulary-learning behavior, we conducted two experiments with university students—the first without speech recognition and the second with speech recognition—and investigated the learning strategy with dictation-type ASR and its superiority from the system logs.

**Keywords**: English vocabulary learning, Automatic Speech Recognition (ASR), learning behavior

## 1 INTRODUCTION

Vocabulary knowledge includes knowing a word's pronunciation as well as its meaning, and improving pronunciation accuracy in English vocabulary learning is important. This study developed a web-based English vocabulary-learning support system utilizing dictation-type automatic speech recognition (dictation-type ASR) (Hirata and Yamada, 2022). Dictation-type ASR transcribes pronunciation and can be freely implemented in a web browser. Many studies have reported that dictation-type ASR enhances pronunciation skills (e.g., Dai and Wu, 2021). However, a specific learning strategy using ASR has not yet been developed. Because learning strategies determine the quality of self-directed learning, a learning strategy using ASR could be useful for improving students' quality of pronunciation learning. Therefore, this study aims to investigate learning strategies with ASR based on the learning logs and compare them with a learning strategy without ASR to reveal the advantages of the tool.

## 2 CASE STUDY

Two experiments (20 minutes in both) were conducted to collect the learning data. In the first experiment, 22 university students used the system without ASR (non-ASR group), and in the second experiment, 25 university students used the system with ASR (ASR group). Subjects in both groups were required to learn 9 words. Lag-sequential analysis, an analysis method that statistically derives the probability that a behavior is likely to follow a certain learning behavior (Bakeman and Gottman, 1997), was conducted to investigate the learning strategies (e.g., Geng and Yamada, 2023).

### 2.1 The system used in this practice

The system investigated in this study consists of an English vocabulary-learning module (see Figure 1) with the Mozilla Developer Network (MDN)'s Web speech API, which transcribes user speech into text. Its recognition rate is 89.4% for native English speakers (Ashwell and Elam, 2017). The system implements an “audio function” to verify the sound of each word from 0.5x to 1x, a “recording function” to record and play back one’s own pronunciation, and a “point function” to explain how to pronounce the word. In the ASR group, the recording function was equipped with Web speech API, and the recognition results were transcribed.

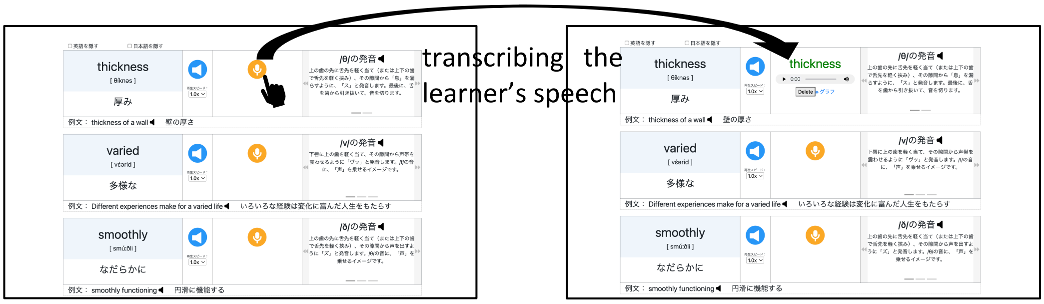



**Figure 1: User interface of the web application**

## 2.2 Learning logs

To validate the learning strategy with ASR, the system learning logs shown in Table 1 were collected, and lag-sequential analysis was conducted.

**Table 1: Learning log**

| Learning log | Explanation |
|---|---|
| Audio | Listen to the audio clip of English vocabulary from 0.5x to 1x |
| Record | Record the pronunciation<br>(ASR group is equipped with Web speech API) |
| Stop_Record | Stop Recording |
| Play_Record | Play back the recording file |
| Open_Point | Open a link that explains how to pronounce the word |

## 2.3 Result of sequential analysis

Total logs collected were 2283 records. The users’ learning strategies are illustrated in Figures 2 and 3. The ASR group showed learning behaviors including Audio 0.8x and Audio 0.5x (Audio→Audio0.8, Audio0.5→Record). Participants may have attempted to listen to the phonemes within the English words in detail by listening at a slower speed. The ASR group also included a sequence of going back and forth between Record and Stop_Record with a higher probability than the non-ASR group (Record⇄Stop_Record). It is assumed that repeated pronunciation and trial-and-error were facilitated by the visualization of speech results using ASR. Furthermore, while the ASR group showed the behavior of opening the pronunciation point after stopping the recording and playing it back (Stop_Record → Open_Point, Play_Record → Open_Point), the non-ASR group opened the pronunciation point in isolation from the other functions. Since “Open_Point” is a function that explains how to pronounce phonemes in a word (e.g., tongue position and breath), the ASR group may have been encouraged to use the function to overcome a pronunciation error. One possible reason for this may be that the transcription of speech results made it easier for participants in the ASR group to recognize their own pronunciation errors and encouraged them to undertake efforts to improve their pronunciation.

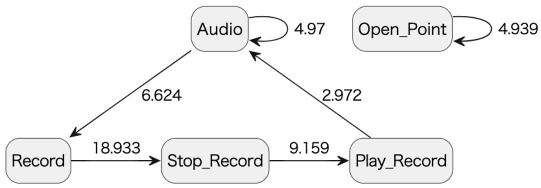



Figure 2: Results of lag-sequential analysis of the non-ASR group (N=22)

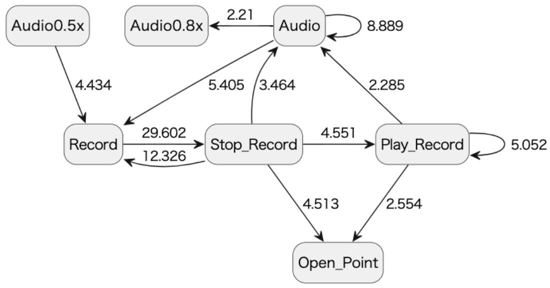



Figure 3: Results of lag-sequential analysis of the ASR group (N=25)

## 3 CONCLUSION AND FUTURE WORK

This study investigated a learning strategy with ASR based on the learning logs of a web system and compared it with a learning strategy without ASR to reveal the advantages of the tool. Two experiments were conducted to collect learning data from the non-ASR and ASR groups. The results of the analysis suggest that ASR promotes repeated pronunciation practice and reflection on pronunciation improvement. Because pronunciation learning using dictation-type speech recognition requires learners to identify pronunciation errors based on dictation texts, learners may reflect on speech form problems, which could contribute to the improvement of their autonomy in utilizing various functions. Future studies should clarify how to present effective scaffolding and recommendations for learners to use these functions to correct their own pronunciation errors.

## ACKNOWLEDGMENT

This work was supported by JSPS Grants-in-Aid for Scientific Research JP22H00552, and JP21KK0184.

## REFERENCES

Ashwell, T. and Elam, J.R., (2017). *How accurately can the Google Web Speech API recognize and transcribe Japanese L2 English learners' oral production? The JALT Call Journal,* Vol. 13, No. 1, pp. 59–76

Bakeman, R., and Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge university press.

Dai, Y. and Wu, Z., (2021). *Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: a mixed-methods study*. *Computer Assisted Language Learning,* Vol. 36, No.1, pp. 1–24.

Geng, X., and Yamada, M. (2022). *Using learning analytics to investigate learning processes and behavioural patterns in an augmented reality language learning environment.* Journal of Computer Assisted Learning, Vol. 39, No. 2, pp.532-546

Hirata, S. and Yamada, M, (2022). *Automatic speech recognition system to enhance the use of vocalization strategy.* Proceedings of 19th International Conference Cognition and Exploratory Learning in Digital Age (CELDA 2022). Lisbon, Portugal, pp. 147–154.

# Towards Creating Feedback Analytics to Support Students' Learning

**Bhagya Maheshi, Hua Jin, Shaveen Singh, Caleb Ooi, Aaron Zheng, Minhua Zhou, Mikaela Milesi, Hiruni Palihena, Roberto Martinez-Maldonado, Yi-Shan Tsai**
Monash University
[Bhagya.GangodaGamachchige, Flora.Jin, Shaveen.Singh, cooi0004, aaron.zheng, minhua.zhou, mikaela.milesi, hiruni.palihena, Roberto.MartinezMaldonado, yi-shan.tsai]@monash.edu

**ABSTRACT**: Feedback is an essential process of learning in higher education. Yet, capturing students' interactions with feedback and understanding the impact of it is challenging. Learning Analytics (LA) can mitigate this challenge by mining large learning datasets produced in technology-enabled learning environments. To address the identified problem, we propose an LA solution - PolyFeed. It is an analytics system that captures and analyses student interactions with feedback, thereby scaffolding students in managing and acting on feedback to improve their learning. We followed a design-based approach, and this paper presents the findings of the prototype phase. Specifically, this study explored students' perspectives on the extent to which PolyFeed's capture and presentation of their interactions with feedback can support learning. The results show that students appreciated the functionalities of organising feedback, formulating actions based on feedback, and PolyFeed aids in tracking and monitoring their progress in implementing feedback. In this paper, we discuss how the analytics about student interactions with feedback (feedback analytics) presented in PolyFeed can potentially enhance learning, specifically through scaffolding feedback literacy and facilitating dialogic feedback.

**Keywords**: Feedback Analytics, Learning Analytics, Feedback Theories, Traceability

## 1 BACKGROUND AND CONTEXT

Feedback is an essential process of learning in higher education. However, monitoring how students engage with, interpret and act on feedback (***traceability***) is challenging (Winstone, 2019). Without understanding how students interact with feedback, educators may struggle to improve their feedback process to support students' learning from feedback. By mining large datasets produced in a technology-enabled learning environment, Learning Analytics (LA) can potentially overcome the challenge of traceability. Yet, the primary aim of existing LA tools has focused on enhancing students' overall experience with feedback by improving feedback content, frequency, and timeliness. However, there is a lack of attention to students' sense-making process of feedback or ways to support students to transform feedback into action (Winstone, 2019). This presents a gap in understanding the effectiveness of feedback, specifically the extent to which students use feedback to improve learning.

Additionally, existing LA-based feedback tools tend to focus on students' behavioural engagement with feedback (e.g., frequency of opening feedback files and duration of student engagement with feedback). The surface-level analytics produced by these tools fail to capture and evaluate the multifaceted impacts of feedback (e.g., cognitive, affective, and relational). In addition, existing LA feedback tools tend to lack alignment with educational theories (i.e., feedback theories) (Tsai, 2022).

To bridge these identified gaps in LA feedback practice, we propose an LA solution - PolyFeed - that builds on feedback theories, including feedback literacy and dialogic feedback (Carless & Boud, 2018). PolyFeed captures and analyses students' interactions with the feedback, defined as Feedback Analytics (FA) in this paper. We followed a design-based approach including seven steps (empathy, define, ideate, prototype, test, involve, and sustain) of a co-design model presented by Carlos et al. (2018). This paper presents the findings of the prototype phase in which we explored students' perspectives on the extent to which PolyFeed's capture and presentation of their interactions with feedback can support learning.

## 2 METHODOLOGY

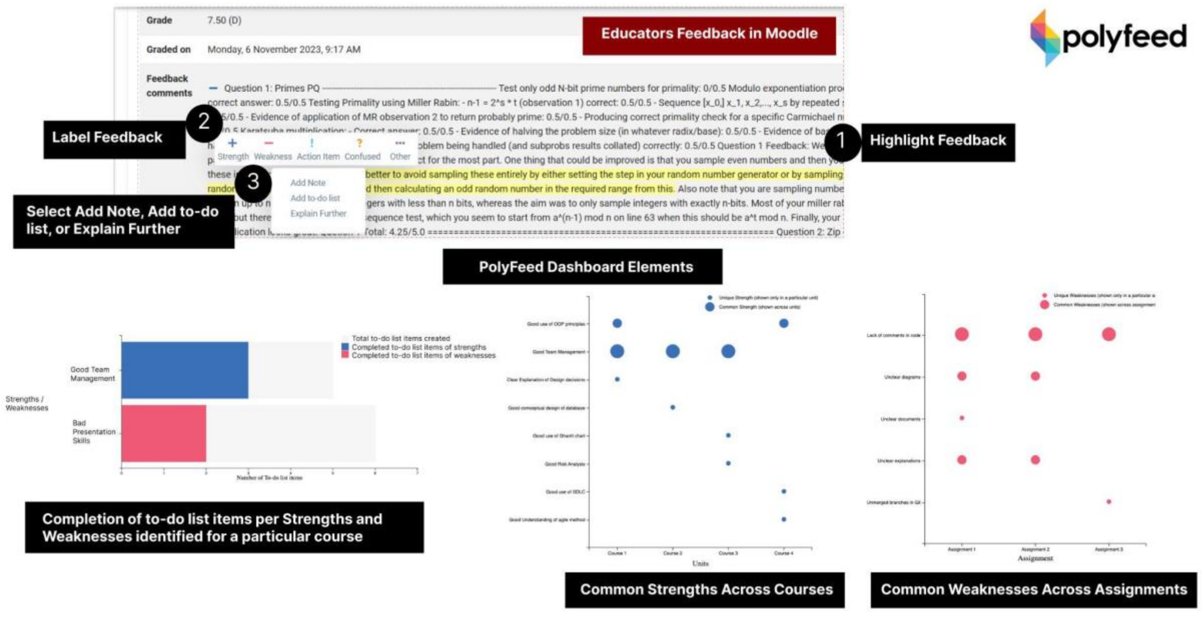



**Figure 1: PolyFeed Features and Dashboard Elements**

**Tool Design:** Based on these findings of focus group discussions with students captured their experience with feedback and guided by feedback literature (Winstone, 2019; Ryan et al., 2021), three primary functions – **Annotate Feedback, Create Work Plan, and View Summary** – were identified as essential features for PolyFeed (See Figure 1). A high-fidelity responsive prototype (https://bit.ly/3ZcDcjK) was developed using Figma.

**Study Design:** We designed four use scenarios of PolyFeed to capture participants' interactions with different modes of digital written feedback and the FA dashboard. Participants were also asked to share their overall perceptions of the tools' potential to support learning. Each validation session lasted an hour. Sixteen students in total participated. The collected data was analysed using a thematic analysis method.

# 3 FINDINGS

The findings show that PolyFeed can scaffold students' feedback literacy (Carless & Boud, 2018), which is key to effective utilisation of feedback. For example, students indicated that PolyFeed helped them perceive feedback as a continuous and valuable source of information by creating a space to revisit and view multiple sources of feedback in one place. In addition, the feature to label the feedback and categorise relevant information as "strength", "weakness", "confusing", "action items", and "other" aids students to make sense of feedback and organise it in a structured way. Notably, students also emphasised that FA has the potential to support them in evaluating and reflecting on learning across courses, identifying areas requiring improvements, and transforming feedback into action. In addition, the validation results show that PolyFeed is capable of facilitating a dialogic feedback process (Carless & Boud, 2018). As an example, students appreciated that PolyFeed can support them in streamlining the feedback clarification process by allowing students to communicate with their educators directly (an option attached with "to-do" lists). This was considered important to support learning as confusion over feedback can impede feedback adoption.

# 4 CONCLUSION AND FUTURE WORK

Student feedback on the PolyFeed prototype shows that PolyFeed can transform the direction of LA-based feedback from unidirectional to bi-direcitional (dialogic), and scaffold student feedback literacy. Our next step is to pilot PolyFeed in real learning settings.

# ACKNOWLEDGEMENTS

This research is funded by Monash University educational innovation funding (The Pitch) and Advancing Women's Success Grant.

# REFERENCES

Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. Assessment & Evaluation in Higher Education, 43(8), 1315-1325.

Carlos G. Prieto-Alvarez, Roberto Martinez-Maldonado, & Theresa Dirndorfer Anderson. (2018). Co-designing learning analytics tools with learners. In *Learning analytics in the classroom* (1st Edition ed., pp. 93–110).

Ryan, T., Henderson, M., Ryan, K., & Kennedy, G. (2021, 4). Identifying the components of effective learner-centred feedback information. *Teaching in Higher Education*, 1–18.

Tsai, Y.-S. (2022, 6). Why feedback literacy matters for learning analytics. In *Proceedings of the 16th international conference of the learning sciences (icls)* (p. 27-34). International Society of the Learning Sciences.

Winstone, N. (2019). Facilitating Students' Use of Feedback: Capturing and Tracking Impact Using Digital Tools. In *The impact of feedback in higher education* (pp. 225–242). Cham: Springer International Publishing.

# An Institutional Approach: Learning Analytics in Programme Evaluation and Improvement

**Author(s): Ada Tse, Chunsang Chan, Albert Chan and Julia Chen**
The Hong Kong Polytechnic University
ada.sk.tse@polyu.edu.hk, chun.sang.chan@polyu.edu.hk, ak.chan@polyu.edu.hk, julia.chen@polyu.edu.hk

**ABSTRACT**: This abstract introduces the Programme Learning Analytics Report (PLAR), a transformative tool implemented in October 2023 across 80 undergraduate programmes. The PLAR aims to empower programme leaders by leveraging learning analytics for the review and enhancement of programme design and curriculum. Through visualisations and comprehensive analysis, the PLAR serves as a key instrument for assessing academic quality and understanding student learning experiences. While effective, it recognises the importance of analysing different student groups and addressing technical challenges. Continuous feedback collection underscores our commitment to refining the PLAR, fostering a culture of learning analytics. Importantly, the design of the next iteration will actively involve programme leaders, ensuring their perspectives shape the tool's evolution. This initiative serves as a catalyst for continuous improvement, aligning our programmes seamlessly with the evolving demands of education and society.

**Keywords**: learning analytics, programme review, evidence-based, institutional

## 1 INTRODUCTION

In higher education, we typically review our courses by gathering feedback from students and faculty, collecting performance data, and having an external evaluator analyse everything to suggest improvements (Saunders, 2011). However, the rise of big data has brought about a shift. Learning analytics (LA) is now crucial in education, especially with the increase in online learning and technology. Some studies show that combining LA with traditional methods provides valuable insights not otherwise obtained, enabling decisions based on concrete data (Armatas & Spratt 2019; Gottipati, & Shankararaman, 2014; Mendez, Ochoa, Chiluiza, & De Wever, 2014; Dunbar, Dingel & Prat-Resina, 2014; Dawson, & Hubball, 2014).

Our university actively promotes the utilisation of LA to pinpoint areas for programme enhancement and elevate the overall quality of student learning. In line with this commitment, a project was initiated in 2021 to design and develop the Programme Learning Analytics Report (PLAR), with the objective of seamlessly integrating LA into our annual programme review. The PLAR serves as a practical solution, providing programme leaders with crucial insights into curriculum effectiveness, with a focus on data-driven assessments of academic standards (the expected level of student attainment) and academic quality (processes facilitating students to meet the established standards for their awards). Uniquely tailored to align with our university's quality assurance measures, this tool enhances efficiency and reduces the time required for data collection in programme reviews.

## 2 THE PROGRAMME LEARNING ANALYTICS REPORT (PLAR)

The Programme Learning Analytics Report (PLAR) relies on a diverse range of data sources, including student records encompassing admission scores, academic performance, students' entry characteristics and other relevant metrics. Additionally, valuable insights are derived from student survey data, including the Student Feedback Questionnaire (SFQ) and surveys focusing on the student learning experience. Collating data from these disparate sources necessitates a meticulous data integration process. The integration is crucial to provide a comprehensive and unified perspective on the programmes at an institutional level. As part of our ongoing efforts to refine the PLAR, we are continuously working on enhancing this data pipeline. The integration process is designed not only to streamline data collection but also to ensure alignment with broader institutional analytics initiatives. This iterative approach allows us to adapt the PLAR to the evolving landscape of institutional data collection and analysis, thereby contributing to a more comprehensive understanding of programme dynamics.

The PLAR integrates diverse tables and visualisations, specifically designed to aid programme leaders, deans, and heads in gaining a nuanced understanding of the students within their programme. Here are some examples:

- A chart traces the average semester GPA of recent cohorts from admission to graduation. This aids in identifying challenging semesters, indicated by lower GPAs across cohorts. Examining this pattern alongside other information, such as subject grades in those semesters, helps pinpoint potential causes. This information is vital for devising follow-up actions to address any issues.

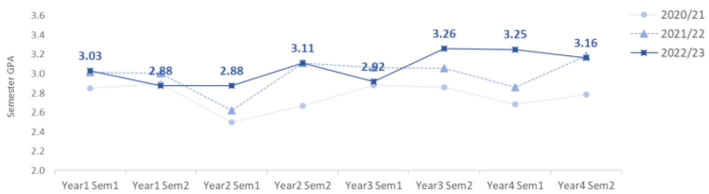



**Figure 1: Trend of semester GPA in each semester for graduates**

- Another chart illustrates the distribution of student grades in specific subjects for each semester in the current academic year. This reveals how graduates within the programme are performing in each subject, highlighting the relative difficulty of subjects. Red arrows draw attention to three subjects with the lowest "B- or above %" or "C- or above %" across all levels, signaling areas that may require attention and potential course improvement measures. This aids readers in assessing current student performance in each subject, understanding the relative difficulty of subjects, and recognising enrollment patterns and subject popularity among students.

| Academic Year | Sem | Subject Code | Subject Level | No. of stud* | Grade distribution A+ A A- B+ B B- C+ C C- D+ D F | No. of graduates achieving each grade A+ | A | A- | B+ | B | B- | C+ | C | C- | D+ | D | F | % of B- or above | % of C- or above |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2022/23 | 1 | ABCD1235 | 1 | 9 | | 0 | 0 | 0 | 0 | 2 | 2 | 1 | 1 | 0 | 2 | 0 | 1 | 44% | 67% |
| 2022/23 | 1 | ABCD1247 | 1 | 28 | | 0 | 0 | 1 | 1 | 8 | 3 | 3 | 1 | 4 | 3 | 1 | 3 | 46% | 75% |
| 2022/23 | 1 | ABCD1254 | 1 | 40 | | 0 | 1 | 2 | 4 | 8 | 6 | 10 | 1 | 4 | 3 | 1 | 0 | 53% | 90% |
| 2022/23 | 1 | ABCD2001 | 2 | 39 | | 0 | 0 | 1 | 2 | 3 | 3 | 9 | 5 | 6 | 2 | 8 | 0 | 23% | 74% |
| 2022/23 | 1 | ABCD2002 | 2 | 29 | | 2 | 4 | 5 | 7 | 4 | 3 | 0 | 2 | 1 | 0 | 0 | 1 | 86% | 97% |
| 2022/23 | 1 | ABCD2003 | 2 | 28 | | 0 | 0 | 1 | 1 | 3 | 4 | 6 | 4 | 3 | 0 | 3 | 3 | 32% | 79% |

**Figure 2: Distribution of subject grades**

# 3 IMPACT

In October 2023, the Programme Learning Analytics Report (PLAR) entered a pilot phase, encompassing all university undergraduate programmes (over 80 in total). Tailored PLARs were distributed to each department, anticipating a significant impact with comprehensive insights into key areas. Subsequent feedback from teachers indicated a high level of satisfaction with the PLAR. They expressed appreciation for the effort and the user-friendly format that consolidates diverse data types into one report, facilitating their preparation of quality assurance reports. Teachers suggested a need to analyze different student groups—graduates, current students, first-year students, and others—allowing for targeted support measures. From a technical perspective, the consolidation of diverse data types remains a challenge. This is due to the varied data structures from which the data originates, compounded by the dynamic nature of programme structures evolving alongside rapid developments in education and society. Looking ahead, involving programme leaders in the next iterations of the tool is recommended to enhance its impact and streamline its utility in their work.

# 4 CONCLUSION

In conclusion, the Programme Learning Analytics Report (PLAR) signifies a pivotal stride towards evidence-based programme enhancement. The forthcoming expansion of the PLAR to include Taught Postgraduate programmes further underscores its transformative potential. As a dynamic tool, the PLAR empowers programme leaders with valuable insights into academic quality and student learning experiences. While currently offering crucial data, the PLAR acknowledges the importance of additional sources for a holistic understanding. The continuous feedback collection reaffirms our commitment to refining this tool, fostering a culture of learning analytics and establishing a systematic approach to evidence-based decision-making in programme review. The PLAR stands as a catalyst for continuous improvement, ensuring our programmes align seamlessly with evolving educational and societal demands.

# REFERENCES

Armatas, C., & Spratt, C. F. (2019). Applying learning analytics to program curriculum review. *The International Journal of Information and Learning Technology*, *36*(3), 243-253.

Dawson, S., & Hubball, H. (2014). Curriculum analytics: Application of social network analysis for improving strategic curriculum decision-making in a research-intensive university. *Teaching and Learning Inquiry*, *2*(2), 59-74.

Dunbar, R. L., Dingel, M. J., & Prat-Resina, X. (2014). Connecting Analytics and Curriculum Design: Process and Outcomes of Building a Tool to Browse Data Relevant to Course Designers. *Journal of Learning Analytics*, *1*(3), 223-243.

Gottipati, S., & Shankararaman, V. (2014). Learning analytics applied to curriculum analysis. *Proceedings of the 2014 AIS SIGED: IAIM International Conference on Information Systems Education and Research*, 2.

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.

Saunders, M. (2011). Programmatic evaluative practice in higher education: design diversity and degrees of programme alignment. In M. Saunders, P. Trowler, P. & V. Bamber, (eds) *Reconceptualising Evaluation in Higher Education. The Practice Turn*. 1st ed. Berkshire, GBR: Society for Research into Higher Education & Open University.

# Eye-Tracking to Study the Effect of Instructor's Presence in Video Lectures: Experimental Design and Data Collection

**Yuli Sutoto Nugroho**
Queen Mary University of London
y.nugroho@qmul.ac.uk

**Marie-Luce Bourguet**
Queen Mary University of London
marie-luce.bourguet@qmul.ac.uk

**Hamit Soyel**
Queen Mary University of London
h.soyel@qmul.ac.uk

**Isabelle Mareschal**
Queen Mary University of London
i.mareschal@qmul.ac.uk

**ABSTRACT**: The purpose of our study is to explore the impact of instructor's presence in video lectures on student's learning gain, visual attention distribution, and cognitive load using eye tracking data. In this paper, we present our experimental design, data collection, and some preliminary results. We compare three experimental conditions: non-lecturer presence in the video, lecturer physical presence, and lecturer presence as an avatar. The data collected comprises pre and post-test knowledge questionnaire results, gaze fixations and pupil diameter from thirty-three study participants. Preliminary results reveal that the lecturer as an avatar condition yields the highest learning gain. An analysis of the number of transitions between the two main areas of interest (i.e., the lecture material and the instructor) indicates intensified students' attention shifts in the physical presence of the instructor. These preliminary findings highlight the impact on learners of the type of instructor's presence in videos, offering insights for instructional design in video-based learning.

**Keywords**: video-based learning, learning gain, visual attention, eye tracking

## 1 INTRODUCTION

Video has emerged as a powerful tool for enhancing learning outcomes and promoting effective information retention. However, it is crucial to understand the impact of various video-based learning (VBL) designs on students' cognitive load, affect and learning gain. Past research, which relies on participants' subjective self-reporting, has shown that online lectures with visuals of the instructor are perceived as more enjoyable. However, the evidence is inconclusive regarding an increase in comprehension (Henderson & Schroeder, 2021). Wang et al. (2017), utilising eye-tracking technology, suggest that the instructor attracts considerable visual attention, and positively influences participants' satisfaction as well as perceived learning.

Our research focuses on examining the effect of different types of instructor's presence on learners' responses using eye-tracking data. We compare three experimental conditions: non-lecturer presence in the video, lecturer physical presence, and lecturer presence as an avatar. We aim to address the following research questions: how does the type of instructor's presence in video affect learning gain (RQ1), learners' visual attention distribution (RQ2), and cognitive load (RQ3)? In this paper, we describe our experimental design, data collection, and some preliminary results.

# 2 METHODOLOGY

We created our video stimuli by adapting a video on the topic of "Deep Learning" from the Lex Fridman YouTube channel (https://www.youtube.com/@lexfridman), dividing it into three segments of approximately 5 minutes each, labelled Topic A, B, and C. The study participants engaged in knowledge tests both before (pre) and after (post) watching the video segments. Each test comprised 5 multiple-choice questions, and the same set of questions was used for both pretest and post-test assessments. During the video-watching sessions, the participants utilised a single camera 120Hz Pupil Core eye-tracker developed by Pupil Labs (https://pupil-labs.com/products/core), with eye-tracker calibration conducted before the viewing session.

Data was collected from 33 participants, all of whom were third-year Electronic Engineering and Computer Science undergraduate students from the same program. This homogeneity in educational background ensured comparable knowledge levels among participants. We randomised the order of the experimental conditions to avoid any habituation effect. The avatar's behaviour is modelled on the lecturer's, although it is simplified (e.g., fewer beat gestures), giving it a quieter presence.

Areas of Interest (AOI) link eye movement measurements to specific sections of stimuli, such as the duration spent focusing on a particular object within the stimulus. We partitioned each video segment into two AOIs: AOI1, representing the lecture material (the presentation slides), and AOI2, denoting the instructor's position. A ratio of 75% for slides on the left side of the video and 25% for the instructor's location on the right side was established. Notably, the right part included a blank space for C1 (no presence), the instructor for C2 (physical presence), and an avatar for C3.

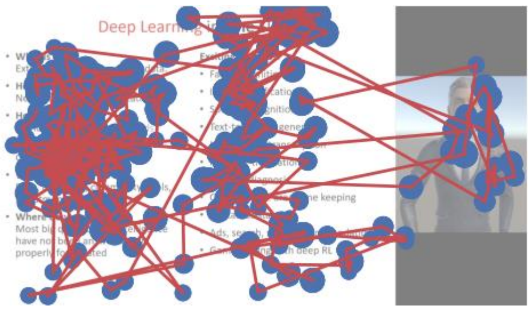


**Figure 1. Example of a scan path in Condition 3 (instructor's presence as an avatar)**

The output of the eye-tracker comprises CSV files that include pupil diameter information, gaze points, and surface fixations. A program was developed using Python code to perform the following tasks: (a) smoothing of the pupil diameter data, filtering out data with low confidence and replacing outliers; (b) calculating gaze fixation durations per AOI, and computing the number of transitions between AOIs; (c) generating visualisations (scan paths and heatmaps) as illustrated in Figure 1.

# 3 PRELIMINARY RESULTS

Comparing learning gains across the three conditions, the initial findings suggest that C3 (presence as an avatar) yields the highest increase in learning gain. A Kolmogorov-Smirnov test revealed non-normal data distribution (Sig < 0.05) for all comparisons, consequently, we applied a Wilcoxon non-parametric test, suitable for non-normally distributed data. The results (Asymp. Sig < 0.05) confirmed significant differences between all conditions (RQ1). The presence of the lecturer as an avatar significantly enhances learning gain, by 46.92% on average using normalised gain (N Gain) calculation as follows: $\%\, N\ gain = \frac{posttest\ score\ -\ pretest\ score}{maximum\ score - pretest\ score}$

In C2 (physical presence), participants exhibited a concentration of 74.51% fixations on slides, with the remaining 25.49% directed towards the lecturer. Conversely, in C3 (avatar), participants allocated 92.10% of their fixation time to the slides, as opposed to 7.99% on the avatar. This difference in the number of fixations on the lecturer AOI between C2 and C3 reveals a stronger visual attraction to the physical lecturer compared to the avatar (RQ2). Examining the number of transitions between the AOIs, where "transition" denotes when students shift their focus between the slides and the lecturer area, and vice versa, reveals that C2 (physical presence) has the most transitions, compared to C3 (avatar) (C1, i.e., no presence, has of course the fewest transitions). 28 people have more transition in C2, 3 people in C3, and 2 people have the same transition number across the two conditions. The physical presence of the lecturer intensifies students' attention shifts between AOIs (RQ2).

Cognitive load indicators, such as pupil dilation measured through eye-tracking technology have been shown to offer insights into the level of cognitive engagement and the mental effort exerted by students (Souchet et al. 2021). 13 people have bigger pupil diameter on C1, 11 people on C2, and 9 people on C3. The variance is very small (var = 0.067684), suggesting similar cognitive load across the three conditions (RQ3). More work is needed in this area, examining relationships between number of transitions, pupil diameter and lecturer versus avatar bodily behaviour. Also, in the future, temporal dynamics will be analysed and pupil data will be normalised for each slide to account for changes in luminance directly affecting pupil diameter. In conclusion, our study's preliminary results suggest that to improve VBL, minimising distracting instructors' behaviours (e.g., beat gestures) may be beneficial to learning. However, to gain a better understanding of the cognitive processes at play, psychological measures such as engagement or learning experience will have to be conducted.

# REFERENCES

Henderson, M.L. & Schroeder, N.L. (2021). A Systematic review of instructor presence in instructional videos: Effects on learning and affect. *Computers and Education Open*, vol. 2. https://doi.org/10.1016/j.caeo.2021.100059

Souchet, A.D., Philippe, S., Lourdeaux, D. & Leroy, L. (2021). Measuring Visual Fatigue and Cognitive Load via Eye Tracking While Learning with Virtual Reality Head-Mounted Displays: A Review. *International Journal of Human-Computer Interaction* 38:9, 801-824. https://doi.org/10.1080/10447318.2021.1976509.

Wang, J. & Antonenko, P.D. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior*, vol. 71, 79-89. https://doi.org/10.1016/j.chb.2017.01.049

# Data Set Size Analysis for Detecting the Urgency of Discussion Forum Posts

**Valdemar Švábenský[1], François Bouchet[1], Francine Tarrazona[2], Michael Lopez II[2], Ryan S. Baker[3]**
[1]Sorbonne University, LIP6 [2]Ateneo de Manila University [3]University of Pennsylvania
valdemar@mail.muni.cz, francois.bouchet@lip6.fr

**ABSTRACT**: In both Massive Open Online Courses (MOOCs) and private courses, instructors face a large amount of queries in discussion forum posts that may merit a response. There has been ongoing research on how to employ machine learning to predict a post's urgency in order to focus instructors' attention. However, it is unclear how large a course is needed to develop these models. We took a publicly available data set of 3,503 labeled forum posts and code from one such prior study. We re-trained the six models described in the study, but with progressively smaller sample sizes, to determine if the models' performance would be preserved. Likewise, we demonstrate that using random subsets even as small as 10% of the original data set achieves comparable performance to full data sets in five out of six models.

**Keywords**: Learning analytics, educational data mining, urgency detection, replication

## 1 INTRODUCTION

When instructors reply to critical forum posts in MOOCs, it may decrease learners' inactivity and dropout rates (Almatrafi et al., 2018, Švábenský et al., 2023). However, it is difficult for instructors to identify which among the sheer volume of discussion forum posts require an urgent response. Thus, it is useful to determine priority posts by utilizing machine learning techniques.

Learning analytics researchers usually attempt to collect as large data sets as possible, but models for predicting post urgency may be useful in smaller courses as well. This paper explores whether models trained on small data sets can achieve comparable performance to using larger data sets.

In the past, few studies have assessed the generalizability of models trained on small data sets of forum posts. Yee et al. (2023) suggest that this approach has potential to be applied across courses in different academic disciplines. Training models on small textual data sets has been explored in other domains, such as urgency detection models in brief crisis messages (Kejriwal & Zhou, 2020). E.g., "Roof collapse in building on Main Street; multiple people trapped inside" is deemed urgent.

## 2 RESEARCH METHODS

We build upon a paper by Švábenský et al. (2023), which evaluated six models trained on a set of 3,503 posts from MOOCs at University of Pennsylvania. The models were validated on a separate test set of 29,604 posts from Stanford University. Post urgency was expressed on a 1 to 7 scale, with 7 as the most urgent. Each forum post text was encoded using *Universal Sentence Encoder v4* numerical feature embeddings. The six best-performing models were: Random Forest (RF), eXtreme Gradient Boosting (XGB), Linear Regressor (LR), Ordinal Ridge Regressor (ORR), Support Vector Regressor (SVR) with a Radial Basis Function kernel, and Neural Network (NN) regressor.

We took the original data and code, which are publicly available (see Švábenský et al., 2023), and implemented a slight modification. We progressively attempted to use data set sizes from 5% to 50% of the original, in increments of 5%. For each data set size, each of the six models was trained ten times, every time on a randomly chosen subset of the given size obtained from the original training set of 3,503 posts. Then, each model was evaluated on the original held out test set of 29,604 posts. Finally, as in the original study, model performance on the test set was assessed using Root Mean Squared Error (RMSE) and Spearman ρ correlation between the predicted and actual values of urgency. The final results were averaged across the ten training runs.

## 3 Results & discussion

As expected, the performance of all models degraded, but to a surprisingly limited extent. Figure 1 shows that across training subsets of different sizes, SVR, NN, and then ORR performed the best, and that a small sample size could be sufficient to train the models. The only exception is the LR model, whose performance degraded substantially until at least 25% of the data set was used. For 5–20%, the average RMSE was as high as 3.60, and the average Spearman correlation only 0.13. From 25% onward, the performance gradually improved with every step of adding more data, and at 50% it reached a satisfactory result of the RMSE of 1.51 and Spearman rho of 0.32.

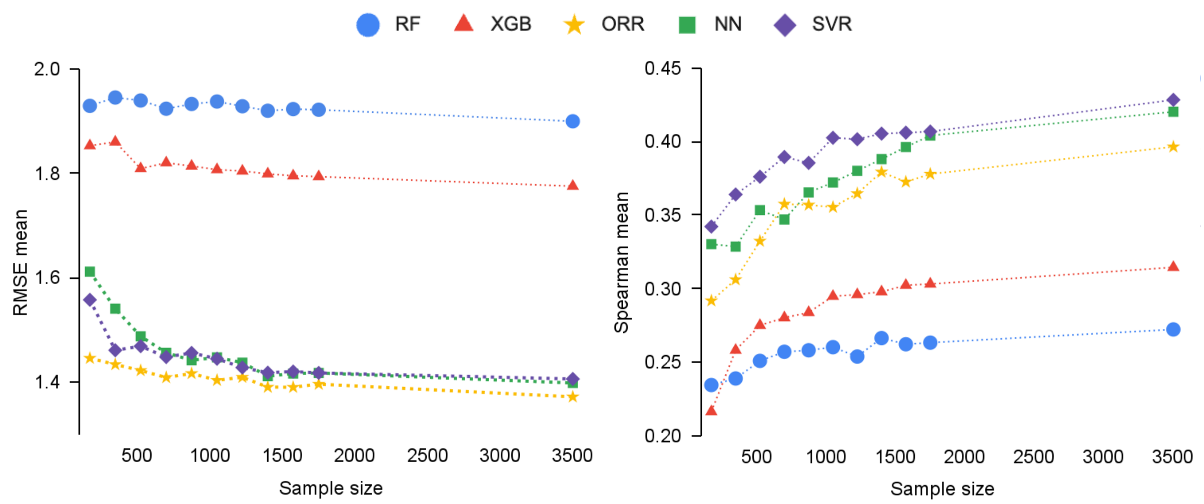



**Figure 1: (left) RMSE and (right) Spearman coefficient with respect to the training data set size. LR is excluded from the figures due to its poor performance on very small samples.**

There is no clear "elbow" in Figure 1 globally across all five models, but for further investigation we arbitrarily selected 10% (350 posts) to illustrate the difference in performance. Table 1 compares the models trained on the (a) original versus (b) partial data set of as little as 350 posts.

**Table 1: Comparison of (a) the results reported by Švábenský et al. (2023), and (b) the same models trained on subsets of 350 posts, evaluated on the original test set, averaged across 10 runs.**

| | (a) Original models | | (b) Models trained on the data subset | |
|---|---|---|---|---|
| Model | RMSE | ρ | RMSE avg, SD | ρ avg, SD |
| RF | 1.8995 | 0.2723 | 1.9499, 0.0530 | 0.2354, 0.0168 |

| | | | | |
|---|---|---|---|---|
| XGB | 1.7753 | 0.3145 | 1.8246, 0.0374 | 0.2633, 0.0165 |
| LR | 1.3953 | 0.3882 | 2.6063, 0.1997 | 0.1575, 0.0343 |
| ORR | 1.3723 | 0.3964 | 1.4349, 0.0306 | 0.3158, 0.0328 |
| NN | 1.3988 | 0.4202 | 1.5376, 0.0513 | 0.3220, 0.0355 |
| SVR | 1.4065 | 0.4283 | 1.4969, 0.0524 | 0.3752, 0.0215 |

## 4 Conclusion and recommendations

The limitation of the past work in this area (Almatrafi et al., 2018, Švábenský et al., 2023) is that it requires time-consuming human labeling of the training data, which not everyone can afford. Identifying the minimal amount of data needed for training prediction models is valuable for replicating this type of detection in other courses and contexts.

Although further research is needed to determine a precise cut-off and demonstrate generalizability, this paper suggests that hundreds, not thousands, of forum posts in the training data set can be sufficient for this problem. Recognizing a point where adding more labeled data does not have a substantial impact anymore can save time for future researchers. Doing so also lowers the barrier of entry to make this approach usable in different contexts, including smaller courses.

Alternatively, future work may utilize weak supervised learning (Zhou, 2018) for detecting urgent posts. Incomplete supervision uses data sets where only a small portion of training data is labeled.

## REFERENCES

Almatrafi, O., Johri, A., & Rangwala, H. (2018). Needle in a haystack: Identifying learner posts that require urgent response in MOOC discussion forums. *Computers & Education*, *118*, 1–9. https://doi.org/10.1016/j.compedu.2017.11.002

Kejriwal, M., & Zhou, P. (2020). On detecting urgency in short crisis messages using minimal supervision and transfer learning. *Social Network Analysis and Mining*, 10(1), 58. https://doi.org/10.1007/s13278-020-00670-7

Švábenský, V., Baker, R. S., Zambrano, A., Zou, Y., and Slater, S. (2023). Towards Generalizable Detection of Urgency of Discussion Forum Posts. In Mingyu Feng, Tanja Käser, and Partha Talukdar (Eds.), *Proceedings of the 16th International Conference on Educational Data Mining* (pp. 302–309). https://doi.org/10.5281/zenodo.8115790

Yee, M., Roy, A., Perdue, M., Cuevas, C., Quigley, K., Bell, A., Ahaan Rungta, & Miyagawa, S. (2023). AI-assisted analysis of content, structure, and sentiment in MOOC discussion forums. *Frontiers in Education*, 8. https://doi.org/10.3389/feduc.2023.1250846

Zhou, Z.-H. (2018). A brief introduction to weakly supervised learning. *National Science Review*, *5*(1), 44–53. https://doi.org/10.1093/nsr/nwx106

# Combination of Computer and Human Analytics for Multimodal Learning Analysis in a Hackathon

**Ayano Ohsaki**
Shinshu University
aohsaki@ohsaki-lab.net

**Daniel Spikol**
University of Copenhagen
ds@di.ku.dk

**Zaibei Li**
University of Copenhagen
zali@di.ku.dk

**ABSTRACT**: This study explores the synergy of computer and human learning analytics in the context of multimodal learning analytics (MMLA). In particular, the study uses observation notes to define key phrases for content analysis using semantic network analysis (SNA), enhancing the qualitative aspect of the analysis. This study analyzed the data from a hackathon supporting women refugees in entrepreneurship based on data from diverse sensors, conversation transcripts, and ethnographic observations. Although our system is a prototype, the study confirms the feasibility of integrating human and computer analyses by visualizing the idea improvement and the participation rate. Future research aims to develop a comprehensive platform considering the temporality of human activities and visualizations for creative endeavors.

**Keywords**: Multimodal learning analytics, Semantic network analysis, Hackathon, Knowledge creation, Human-computer learning analytics

## 1 BACKGROUND

Over the last decade, multimodal learning analytics (MMLA) has begun to offer researchers new tools to bring together diverse streams of sensors with human observations. Researchers use these multimodal systems in real-world learning settings (Martinez-Maldonado et al., 2023). Regarding the analysis of collaborative creative work for learning, socio-semantic network analysis (SSNA) has been known for knowledge creation (Chen et al. 2022). According to Chen et al. (2022), the analysis approach for collaborative discourse has three domains: the cognitive domain is revealed by analyzing content; the social domain is revealed by analyzing group dynamics; and the integrative domain is indicated by synthesizing the social and cognitive domains—moreover, MMLA research on adding content analysis to group dynamics analysis. Therefore, developing a new analytical approach to capture learning from content analysis and group dynamics with sensor data is necessary.

This study proposes the combination of human and computer learning analytics based on observation data, conversation transcript, and voice data. In our proposed approach, we used semantic network analysis (SNA), which focuses on the content aspect of SSNA, and speaker diarization to analyze data from collaborative activities in a hackathon. As a novelty, we used observation notes to define key phrases. Key phrase definition is critical to SNA, and previous research used ways such as picking terms from educational materials by teachers or ranking terms of high frequency by computational techniques (Ohsaki & Oshima, 2019; Chen et al., 2022). However, the former has difficulty capturing “new” ideas in creative activities, while the latter has difficulty capturing changes in ideas that cannot be expressed in numerical frequencies. Hence, we use observation notes to define key phrases. We hypothesized that this would allow for a qualitative perspective to be added to the analysis and that using short but observational records would better interpret the visualized results. Our research question is, what can be described by combining multimodal from different sensors, including automatic capture and transcription of conversations combined with ethnographic observations?

## 2 METHODS

The case for the research is a hackathon organized by different NGOs to support empowering women refugees in entrepreneurship (Kuckertz et al., 2023). We researched the first part of a series of events focused on the Business Canvas Model creation, team formation, ideation, concept refinement, and first-round presentations. Data was collected from a group in the hackathon, which was held to foster the development of business ideas through teamwork and mentorship. This study chose a group of four people who consented to join the study. The hackathon consisted of five sessions with different activities and times. For example, in Session 1 (48-min, 277 lines), all members discussed a café project and refined their understanding using the 5 Whys technique based on a mentor's guide. In Session 2 (56-min, 279 lines), the members navigated the café project, and a member shared experiences from her mother's café, and mentor insights refined the team's ideas. We used different sensors that collected members' positions and captured audio with on-fly transcription and post-priori speaker diarization. Additionally, the workshops were observed by a trained ethnographic researcher whose notes became part of the data stream.

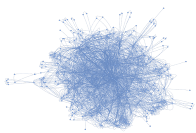
We aim to construct an analytics platform for multimodal analysis, integrating both computer and human analysis. This study represents the inaugural stage of this platform's development, delving into its potential advantages and challenges. Our analysis system for this study was a prototype that consisted of several existing AI tools: an audio data analysis pipeline (Ouhaichi et al., 2021), transcribing audio data (transcriptor, n.d.), translating from Russian and Ukrainian to English (DeepL, n.d.), and summarizing observation notes (Open AI, n.d.). Besides, we chose all nouns and proper nouns from observation notes except the person's name as key phrases using spaCy (Explosion, n.d.). For example, the number of key phrases were 76 in Session 1 and 81 in Session 2.

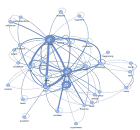
## 3 RESULTS AND FUTURE WORK

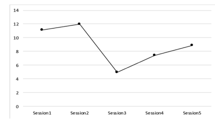
First, Figs. 1a and 1b show the difference between the SNAs using the observation notes. Figure 1a constructs a network encompassing all nouns and proper nouns from Session 1, providing insight into idea changes beyond frequency rankings. However, with 463 key phrases, this network requires simplification for enhanced interpretability. Conversely, Fig. 1b extracts 76 key phrases from observation notes, revealing a network where members emphasize terms like “problem” and “people,” highlighting a focus on addressing people's problems as an initial step in business plan creation. The core idea of their business plan, “food,” was already discussed in this session.

Figures 1c and 1d show the graphs for evaluating each session regarding idea improvement and participation rate. Following the previous study (Ohsaki & Oshima, 2019), which expressed changes in ideas as the sum of the degree centrality of key phrases, Fig. 1c shows the sum of degree centrality at the end of the sessions. On the other hand, Fig. 1d uses speaker diarization techniques to show the silence ratio and the participation equality. These graphs show that Session 2 had a higher participant equality and a lower silence ratio than Session 1, indicating that the participants considered various ideas using a wide range of key phrases in Session 2. Since Session 3 had a higher silence ratio, it is unsurprising that the content analysis results were lower than those of the early phase. Although the total degree centrality score in Fig. 1c is lower in Session 4 than in Sessions 1 and 2 because the ideas converge toward the final presentation, the graph in Fig. 1d shows that all members participated in the discussion.

Although this study is a prototyping study, we have confirmed that combining human analysis in observation notes and computer analysis on conversation and voice data can represent the state of collaborative creative activity. Future research will include the development of an integrated platform, developing an analysis method that considers the temporality of human activities, and visualizations of the activity process based on our proposed method. For example, we illustrate how the mentor's guidance affects the participant's activity.


a. SNA of the standard analysis


b. SNA of the proposed approach


c. total degree centrality of each session

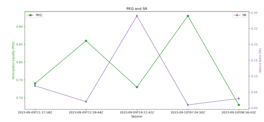

d. participant rate of each session

**Figure 1: result of visualizations.**

## ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Numbers 18K13238, 22H01043 and 23K11357. Additional funding from the German Federal Government Commissioner for Culture and the Media with funding from the European Commission.

## REFERENCES

Chen, B., Zhu, X., & Shui, H. (2022). Socio-semantic network motifs framework for discourse analysis. *LAK22: 12th International Learning Analytics and Knowledge Conference* (pp. 500-506).

Explosion (n.d.). spaCy. Retrieved November 23, 2023, from https://spacy.io/

Martinez-Maldonado, R., Echeverria, V., Fernandez-Nieto, G., Yan, L., Zhao, L., Alfredo, R., Li, X., Dix, S., Jaggard, H., Wotherspoon, R., Osborne, A., Shum, S. B., & Gašević, D. (2023). Lessons learnt from a multimodal learning analytics deployment in-the-wild. *ACM Transactions on Computer-Human Interaction*, 31(1), 1–41.

DeepL (n.d.). Retrieved December 12, 2023, from https://www.deepl.com/translator

Kuckertz, A., Bernhard, A., Berger, E. S. C., Dvouletý, O., Harms, R., Jack, S., & Kibler, E. (2023). Scaling the right answers – creating and maintaining hope through social entrepreneurship in light of humanitarian crises. *Journal of Business Venturing Insights*, 19.

Ohsaki, A., & Oshima, J. (2019). An Assessment of Learning as Knowledge Creation by Using Socio-Semantic Network Analysis and Text Mining, *Japan Journal of Educational Technology*, 43(1), 13-29.

Open AI (n.d.). ChatGPT. Retrieved November 30, 2023, from https://chat.openai.com

Ouhaichi, H., Spikol, D., & Vogel, B. (2021). Mbox: Designing a flexible iot multimodal learning analytics system. *2021 International Conference on Advanced Learning Technologies (ICALT)* (pp. 122-126). IEEE.

Transkriptor (n.d.). Retrieved December 12, 2023, from https://transkriptor.com/

# mBox-audio: Unveiling Conversational Dynamics through Real-Time and Post-Time Audio Analysis for MMLA

**Authors: Zaibei Li, Daniel Spikol, Liv Nøhr**
University of Copenhagen
zali@di.ku.dk, ds@di.ku.dk, lsnl@ind.ku.dk

**ABSTRACT**: Multimodal Learning Analytics (MMLA) has emerged as a powerful tool for understanding and improving learning outcomes. However, capturing and analyzing conversational patterns in educational settings poses challenges due to privacy concerns, performance limitations, and the complexity of classroom environments. To address these challenges, we present mBox-audio, a new system that provides comprehensive insights into conversational dynamics through real-time and post-time audio analysis. The system utilizes two distinct audio pipelines: a real-time analyzer for monitoring conversations and a post-time analyzer for retrospective analysis. Preliminary results demonstrate the effectiveness of mBox-audio in speaker diarization and speech transcription, with comparable performance to commercially available transcribers. Future plans include fine-tuning the Whisper model to enhance transcription accuracy for specific tasks and further develop the real-time analysis feature to fully utilize its potential for enhancing MMLA in educational settings.

**Keywords**: Multimodal Learning Analytics, Speaker Recognition, Automatic Speech Recognition

## INTRODUCTION

Conversation, a cornerstone of interpersonal communication and teamwork, is particularly pivotal in educational settings that facilitate learning and collaboration. However, effectively analyzing these interactions in classrooms presents unique challenges. Recent studies have sought to address these challenges: Emily et al. (2020) developed an automated feedback system using wireless headsets and machine learning for discourse analysis in teaching environments. Canovas and Garcia (2022) employed speaker diarization with SVMs to categorize classroom audio into teacher, student, and silence segments. Despite these advancements, these methodologies predominantly process data post-activity and often focus narrowly on specific aspects of classroom discourse. This underscores the need for systems capable of real-time analysis and a more comprehensive understanding of conversational dynamics in educational contexts.

Our research introduces the "mBox-audio" to address an innovative IoT audio data pipeline in this unexplored application space. Designed for educational contexts focused on group work, it integrates data collection, speaker recognition, and speech recognition to monitor and analyze conversations in real time and retrospectively. This approach facilitates a deeper exploration of conversational characteristics within educational settings. Our central research question is, "*How can audio collection and analysis be optimized for Multimodal Learning Analytics (MMLA)*?"

The following sections detail the development of two distinct audio pipelines within the "mBox-audio" system. We present initial findings on conversational characteristics extracted from our system and assess its accuracy in speaker diarization and speech transcription. Finally, we conclude with a discussion on the preliminary results.

# DESIGN AND IMPLEMENTATION

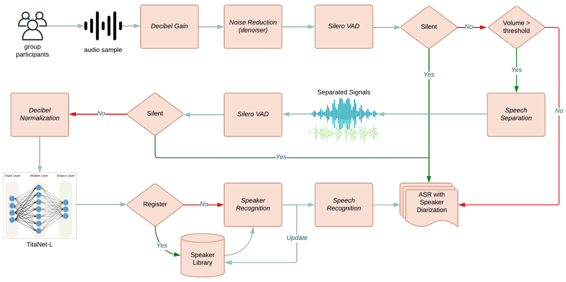



**Figure 1: Real-time Audio Analyzer**

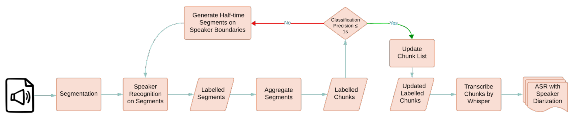



**Figure 2: Post-time Audio Analyzer**

Figure 1 illustrates the real-time audio analyzer pipeline within the mBox-audio system. This pipeline processes consecutive, fixed-duration audio samples sourced from single or multiple connected recorders. The audio samples are subjected to preliminary processing, including decibel normalization, noise reduction, voice activity detection, and speech separation. Subsequently, the TitaNet-L (Koluguri et al., 2021) model is employed to generate speaker embeddings. These are compared with registered speaker embeddings to identify the speaker with the highest similarity that exceeds a set threshold. For each speaker turn, the Whisper-large-v3 model (Radford et al., 2022) transcribes the speech chunk. The duration of the audio segments is adjustable; while shorter durations enhance the granularity of speaker recognition, they may compromise the accuracy of both speaker recognition and transcription as a trade-off.

Figure 2 shows the post-time audio analyzer pipeline which begins by ingesting an entire audio file, subsequently segmenting it into portions, and then identifying and labelling speaker turns for each segment. Next, targeting the boundaries of the identified speaker turns, the system performs a half-scale segmentation and speaker recognition. This multi-scaled segmentation approach recursively refines the precision of speaker diarization. After achieving the targeted precision through iterative refinement of the labelled segments, the audio is transcribed sequentially, chunk by chunk, to ensure transcription accuracy aligned with the speaker turns.

# PRELIMINARY RESULTS

**Table 1: Preliminary comparative analysis with manually annotated transcription**

| Metrics | Ground and Otter AI | Ground and mBox-audio |
|---|---|---|
| Diarization Error Rate (DER) | 23.2% | 18.5% |
| Word Error Rate (WER) | 36.3% | 27.9% |
| Match Error Rate (MER) | 35.8% | 27.3% |
| Word Coverage Rate (WCR) | 81.6% | 85.4% |

In our preliminary experiment, we assessed the diarization and transcription capabilities of the mBox audio system on a 10-minute audio clip from a one-hour noisy meeting recording, captured by a Jabra Speak2 75 device. This clip was transcribed for comparative analysis using Otter AI (otter.ai) industrial transcriber, with manually annotated speaker identities. The mBox-audio system's post-time analyzer, with a pre-registered speaker library set to segment every 4 seconds, processed the same audio clip with an intel i7-12700H processor in 25 minutes. Ground truth speaker transcription logs were manually generated for this clip. Figure 3 presents an interactive visualization graph produced by the system, illustrating session-level conversational characteristics, including average pause duration, normalized turn-taking count, participation equality, and silence ratio. Table 1 compares the diarization error rate, word error rate, match error rate, and word coverage rate among Otter AI, mBox-audio, and manual annotations. Figure 4 displays the visualized Rich Transcription Time Marked files, sequentially showing the outputs from Otter AI, mBox-audio, and the ground truth transcriptions.

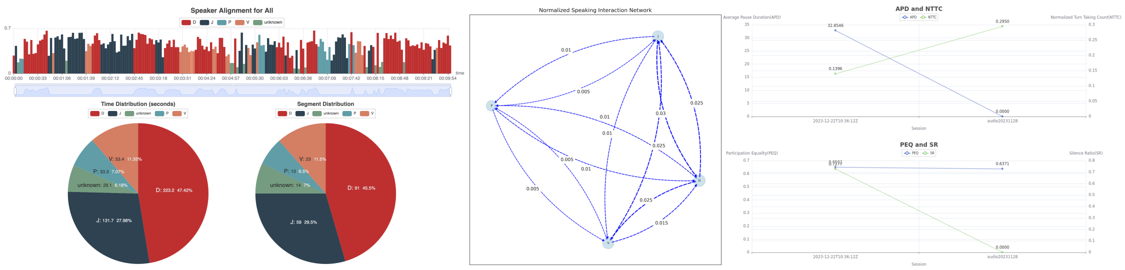



**FIGURE 3: INTERACTIVE SPEAKER DIARIZATION GRAPH AND POST-ANALYSIS FROM MBOX-AUDIO**

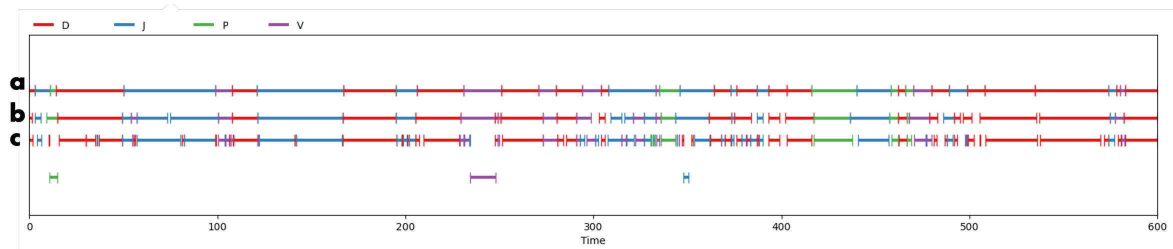



**FIGURE 4: SPEAKER DIARIZATION GRAPH GENERATED WITH PYANNOTE (PLAQUET & BREDIN, 2023) FROM OTTER AI (A), MBOX-AUDIO (B), AND MANUALLY ANNOTATED TRANSCRIPTION (C)**

## DISCUSSION

While our post-time analyzer performs similarly to commercial transcribers, it has difficulties analyzing short comments common for group work. In optimizing audio collection and analysis for group work, we need to compromise between the granularity of speaker recognition and accuracy. Furthermore, choosing to work in real time for sensitivity reasons requires retaining audio data only in the processing phase, which in turn adds more importance to choosing the right duration of audio segments for both encapsulating the micro-communication common for groups and being able to recognize speakers. Thus, these strategies also optimize the audio collection and analysis for MMLA.

## REFERENCES

*Canovas, O., & Garcia, F. J. (2023). Analysis of Classroom Interaction Using Speaker Diarization and Discourse Features from Audio Recordings. In M. E. Auer et al. (Eds.), ICL 2022, LNNS 634 (pp. 67–74). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-26190-9_7*

*Jensen, E., Dale, M., Donnelly, P. J., Stone, C., Kelly, S., Godley, A., & D'Mello, S. K. (2020). Toward Automated Feedback on Teacher Discourse to Enhance Teacher Learning. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20), April 25–30, 2020, Honolulu, HI, USA. ACM. https://doi.org/10.1145/3313831.3376418*

*Koluguri, N. R., Park, T., & Ginsburg, B. (2021). TITANET: Neural Model for Speaker Representation with 1D Depth-Wise Separable Convolutions and Global Context. arXiv:2110.04410v1.*

*Plaquet, A., & Bredin, H. (2023). Powerset multi-class cross entropy loss for neural speaker diarization. In Proceedings of Interspeech 2023, August 20-24, Dublin, Ireland.*

*Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). Robust Speech Recognition via Large-Scale Weak Supervision. arXiv preprint arXiv:2212.04356.*

# A nationwide investigation of a technology-based formative assessment tool that provides analytic data for teachers

**Linlin Li, Natalie Brezack, & Mingyu Feng**
WestEd
lli@wested.org, nbrezac@wested.org, mfeng@wested.org

**ABSTRACT**: This poster presents early findings from a nationwide study of a professional learning community-augmented independent practice and formative assessment educational technology platform, ASSISTments. In ASSISTments, teachers assign independent practice work and students complete work while receiving feedback and hints. Teachers then receive key analytic data about students' performance that they can review with students and use to adapt their classroom instruction to their students' needs. We present results from the first cohort of a quasi-experimental design study on middle school students' math achievement and a case that demonstrates teachers' use of ASSISTments analytic data in classrooms.

**Keywords**: data-driven instruction, formative assessment, math achievement

## 1 INTRODUCTION

Research has shown that regular independent practice and teachers' use of resulting data to inform instruction are crucial for student learning (Irons & Elkington, 2021). However, traditional, paper-based independent practice has limitations: when students complete their work, they may make errors and practice incorrectly, and teachers have limited time to analyze students' progress, adjust their subsequent instruction, and determine students' learning needs (Mendicino, et al., 2009). The ASSISTments platform provides opportunities to address these challenges by providing independent practice for students and analytics of formative assessment data for teachers.

Developed by researchers at Worcester Polytechnic Institute, the ASSISTments platform aligns with the theoretically- and empirically-based instructional practices of formative assessment (Heritage & Popham, 2013). ASSISTments follows four steps of formative assessments. Step 1: Teachers create math assignments aligned with grade-level standards. Step 2: While completing assignments independently, students receive immediate feedback, hints, and explanations in ASSISTments to support their understanding and problem-solving. Step 3: Teachers obtain real-time assignment reports that summarize individual and class performance and common wrong answers, and gain insights on students' needs for support. Step 4: Teachers analyze the data with the class and discuss common errors, using the information from the reports to inform their subsequent instruction and meet student needs. The ASSISTments platform has been used in middle schools since 2004 and has been found to improve students' math learning outcomes in numerous rigorous studies (Roschelle et al., 2016; Sahni et al., 2021), including a study concluded during COVID-19 (Feng et al., 2023).

## 2 PRESENT STUDY

The purpose of this study was to understand the impact of a virtual professional learning community (vPLC)-augmented ASSISTments intervention on students' math achievement across 12 states in the United States. A vPLC was developed to facilitate training and engage teachers around the nation in discussions. Teachers participated in initial trainings and regularly monthly meetings led by vPLC facilitators to guide teachers to use the four steps of formative assessment with ASSISTments. Sixteen teachers and 717 students from 14 schools in 12 states participated in the first cohort of the study during the 2022-23 school year. A second cohort of 49 teachers from 30 schools in 22 states is participating during the 2023-24 school year.

This study used a quasi-experimental design: Students using ASSISTments were compared with a virtual comparison group drawn from a national testing database that was matched with the intervention students on beginning-of-year pre-test performance and demographic background information. Students' math achievement was measured at the start and end of the school year with the online MAP Growth assessment provided by NWEA (https://www.nwea.org/map-growth/). The teachers participated in an interview during which they were asked about their implementation of ASSISTments' four steps of formative assessment. A subset of teachers was observed during classroom observations and participated in follow-up interviews.

## 3 RESULTS

The study used a partially-nested hierarchical linear model with intervention students nested in classrooms to analyze the impact of the vPLC-augmented ASSISTments intervention on students' math performance. The results indicated that the first cohort of participants outperformed the virtual comparison group as measured by NWEA MAP, although the difference between groups was not statistically significant. Within the intervention group, 60.68% of students who used ASSISTments maintained or improved their math performance to above the 50th percentile on the MAP national benchmark assessment. 61.03% of low-performing students (whose baseline math performance was below 50th percentile) improved to above the 50th percentile.

We noticed that teachers who encouraged their students to complete their ASSISTments assignments and reviewed ASSISTments reports more frequently tended to have students with better performance on MAP. For example, at one school with high student MAP performance, 82.3% of students completed ASSISTments assignments (whereas on average, 70.7% of students at other schools completed assignments). Likewise, students at this school completed on average 82.3% of problems in ASSISTments assignments, compared with an average problem completion rate of 72.3%. This high assignment and problem completion rate made ASSISTments' analytics more informative for teachers since the reports pertained to more students and covered more problems. The teachers at this school used the rich analytic data from ASSISTments to guide their instruction. One teacher provided an example of their reaction to looking at ASSISTments' data (Figure 1): "*You look at ASSISTments [data]... oh my gosh, everyone missed number four. We gotta go over number four tomorrow.*" Another teacher reported that the ASSISTments data was valuable, surprising, and guided instructional decision-making: "*a couple times where I'm like, okay, I think everyone's getting it... It feels like, you know, it's a rockstar day… And I'd look at the data from ASSISTments. It's like, okay, [the students] bombed it… let's slow down... they weren't quite getting it as well as I thought they were.*" During classroom observations, this teacher projected ASSISTments data while students were working to draw their attention to a common error on one problem and show how well the class was performing on another problem (similar to Figure 1).

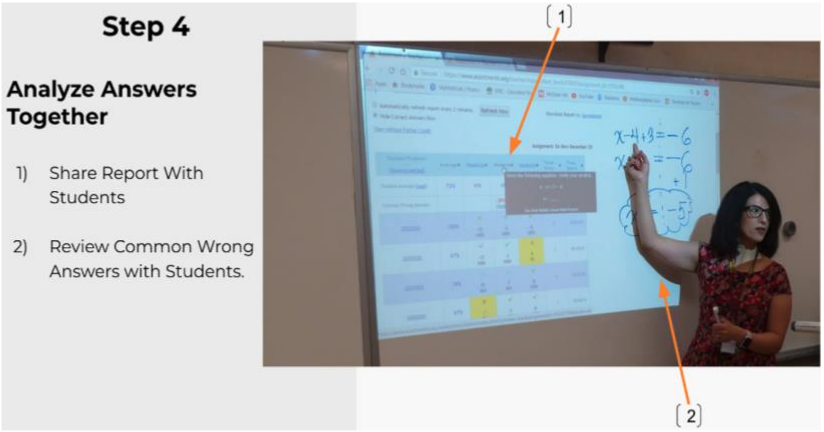



Figure 1: A teacher engaging in Step 4 of formative assessment with ASSISTments by displaying student analytic data on performance and common wrong answers to inform a class discussion.

## 4 DISCUSSION

Findings from this nationwide vPLC-augmented ASSISTments intervention study indicated that using analytic data from a formative assessment educational technology platform could inform teachers' instruction and further students' math learning. When teachers encourage their students to complete their work in educational technology platforms, they can use the resulting rich analytic data about students' progress and gaps in their understanding to inform instruction. Future research should continue to explore opportunities for technology-based programs to provide analytics for teachers to inform data-driven instruction and encourage students' active learning.

## REFERENCES

Feng, M., Huang, C., & Collins, K. (2023). Promising Long Term Effects of ASSISTments Online Math Homework Support. In *Proceedings of International Conference on Artificial Intelligence in Education, A Late Breaking Result*. pp. 212-217. Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36336-8_32

Heritage, M., & Popham, W. J. (2013). *Formative assessment in practice: A process of inquiry and action.* Cambridge, MA: Harvard Education Press.

Irons, A., & Elkington, S. (2021). *Enhancing learning through formative assessment and feedback*. Routledge.

Mendicino, M., Razzaq, L., & Heffernan, N. (2009). A comparison of traditional homework to computer-supported homework. *Journal of Research on Technology in Education, 41*(3), 331.

Sahni, S. D., Polanin, J. R., Zhang, Q., Michaelson, L. E., Caverly, S., Polese, M. L., & Yang, J. (2021). *A what works clearinghouse rapid evidence review of distance learning programs.* US Department of Education.

# Unveiling Student Interests: A Topic Modeling Exploration of Interest-Driven Project-Based Learning

**Michael Smith, Vishesh Kumar, Marcelo Worsley**
Northwestern University
mssmith@u.northwestern.edu, visheshkay@gmail.com,
marcelo.worsley@northwestern.edu

**ABSTRACT**: In this poster, we analyze a cross-disciplinary computer science project course as a case study for mapping students' exposure across different topics using topic modeling analyses on the students' project proposals. We find student interests to vary across aspects such as learning modality and means of data collection, resulting in a mapping hierarchy of nuanced relationships. This mapping offers a design heuristic for using written plan submissions in project-based learning courses to identify student learning, interests, and grounds for revising a course's design during its conduct and between iterations.

**Keywords:** topic modeling, project-based learning, computer science education

## 1 INTRODUCTION

Recognizing student learning from project-based learning (PBL) environments is a significant challenge (Guo et al., 2020). While multimodal learning analytics approaches have highlighted initial lenses into highly detailed process-oriented learning interactions in PBL environments (Worsley et al., 2016), they often rely on equipping spaces and youth with advanced technologies that are often expensive and challenging (Kajamaa & Kumpulainen, 2020). Topic modeling, specifically Latent Dirichlet Allocation (LDA), has been the leading contemporary method to conduct accessible semantic analyses on a variety of learner texts, with many efforts focusing on students' essays. However, few have used topic modeling in project-based learning, such as Fwa (2021), who modeled project reflections to inform future course revisions.

We build on their approach by expanding our analysis beyond word clouds to develop conceptual relationship mappings between topics to highlight an overlooked space for using well-established Learning Analytics methods for understanding project-based learning experiences. Students' written project plans and summaries can offer rich insights into their interests. These can help educators gain insights into their students' trajectories and how to support them in their plans during the course duration. In this study, we demonstrate using a topic modeling analysis on the students' project proposals to answer the questions: Q1. Which technology skills or interest areas are students drawn to and leverage in their final projects? Q2. How do the skills or interest areas relate to each other? This usage of topic modeling demonstrates an analytical method that provides educators with a powerful example of leveraging Learning Analytics methods in project-based courses and cross-interest learning environments.

## 2 METHODS

We examined three iterations of a Computer Science project course during the 2020 and 2022 academic years where the enrollment had some representation from all years but was predominantly senior undergraduates and master students. The aspects of the course were structured to help them learn skills relevant to making a final project of their choice among the following: creating a custom wearable device, using data analytics to model phenomena, or designing a learning experience curriculum. For our questions, we analyze the final project proposals that give insight into the chosen problem contexts and solution methodologies the students took up. We used LDA topic modeling to develop an overall mapping of project topics. These topics are helpful in getting a sense of the diversity in students' interests and discovering with which topics a project report is most closely associated. We use the findings from this to group reports for further analysis.

## 3 FINDINGS & DISCUSSION

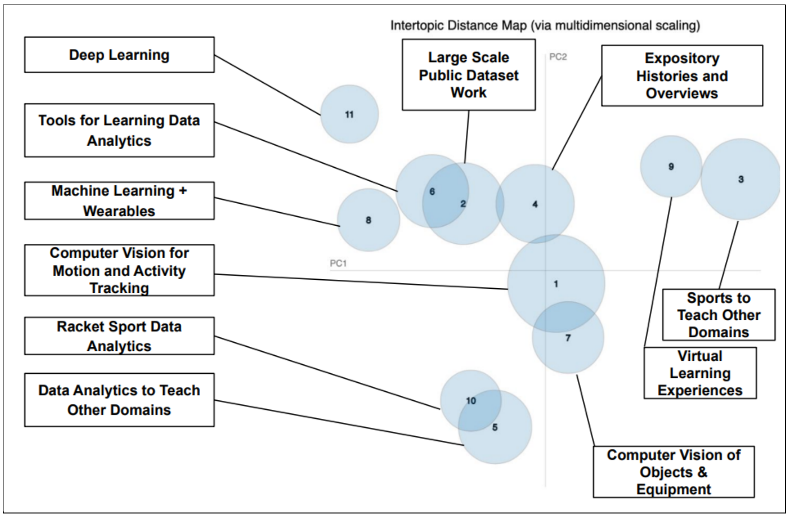



**Figure 1. Intertopic distance map with topic labels**

### 3.1 Interest Areas.

From the model, we found topics that demonstrate the specific skills and interest areas students took up for their projects. Figure 1 shows an intertopic distance map that spatially indicates how distinct these topics are from each other and shows a descriptive label for each topic. No topics are wholly contained within others, indicating that each topic is distinct, though some may be closely related to others. We next categorized the reports, generating a topic distribution for each and labeling them based on which topic had the highest probability. Looking at the papers with the highest probability for each topic allowed us to generate a more interpretable label for each topic.

### 3.2 Relationships

Starting with the named topics, we contextualized them and their relationships to each other and developed a hierarchy for understanding what it tells us about where interests lie in the class projects. In Figure 2, the larger boxes represent topics, and the smaller boxes are additional inferred contexts. Under each required project area, we see branching relationships based on interest, modality, means, and subsets based on the language used. Ultimately, a chart like this will grow and change with each iteration of the course, becoming more valuable as a tool to see where past interests lay, to follow the development of new interests, and to help surface gaps to be included in future iterations.

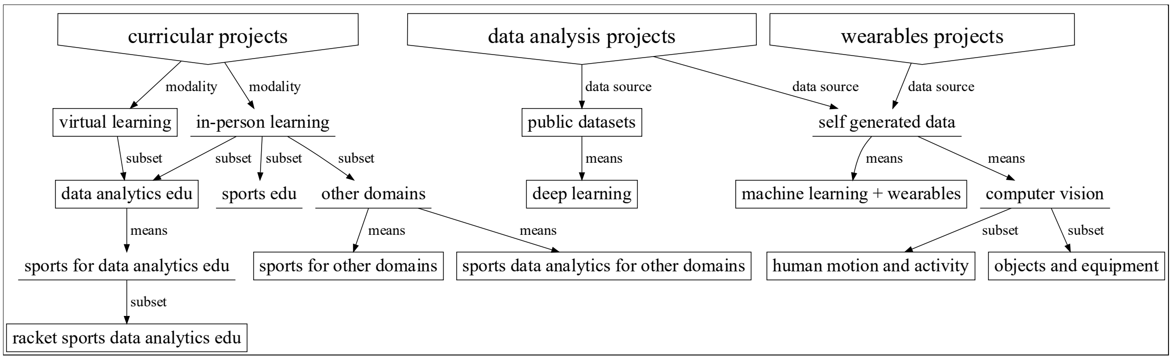



**Figure 2. A relationship mapping of topics with added context**

## 4 CONCLUSION

The benefits of topic modeling in this context are twofold; first, it offers a quick way of linguistically distinguishing word usage patterns that we can correlate to topics of interest to our students and categorize their documents into those topics. Second, with additional context, we can generate robust conceptual hierarchies of how those interest areas relate to each other and see distinctions that influence where those interests differentiate, such as a choice of tool, data source, or modality. The generation of this mapping can allow course designers to better prepare both for the topics discovered and the gaps by understanding how these topics relate to each other.

We hope this work provides a powerful and easy-to-implement example of implementing topic modeling on project-based learning submissions to help instructors and researchers recognize student interests at a classroom level and develop conceptual tools using these insights.

## REFERENCES

Fwa, H. L. (2021). Enhancing Project Based Learning with Unsupervised Learning of Project Reflections. 2021 5th International Conference on Digital Technology in Education, 117–123. https://doi.org/10.1145/3488466.3488480

Guo, P., Saab, N., Post, L. S., & Admiraal, W. (2020). A review of project-based learning in higher education: Student outcomes and measures. International Journal of Educational Research, 102, 101586. https://doi.org/10.1016/j.ijer.2020.101586

Kajamaa, A., & Kumpulainen, K. (2020). Students' multimodal knowledge practices in a makerspace learning environment. International Journal of Computer-Supported Collaborative Learning, 15(4), 411–444. https://doi.org/10.1007/s11412-020-09337-z

Worsley, M., Abrahamson, D., Blikstein, P., Grover, S., Schneider, B., & Tissenbaum, M. (2016). Situating multimodal learning analytics. Proceedings of International Conference of the Learning Sciences, ICLS, 2, 1346–1349.

# Example of a Participatory Approach in Developing Analytics Work at a Science-focused Teaching and Learning Centre

**Zohreh Moradi**
University of British Columbia
moradi@science.ubc.ca

**Noureddine Elouazizi**
University of British Columbia
noureddine.elouazizi@science.ubc.ca

**Warren Code**
University of British Columbia
warcode@science.ubc.ca

**Gülnur Birol**
University of British Columbia
birol@science.ubc.ca

**ABSTRACT**: This poster outlines the process of employing a participatory approach to develop learning analytics goals and processes within the Faculty of Science at the University of British Columbia (UBC). Through collaborative efforts with faculty members and departments, three successive needs analyses, in the period from 2015 to 2022, were conducted to identify opportunities for making use of learning and academic analytics to foster students' success. Interacting with our departments and programs through meetings, consultations and a workshop we facilitated a participatory approach that fostered engagement. This approach revealed commonalities, identified data availability gaps and increased buy-in, strengthening the foundation for our future work in this domain.

**Keywords**: Participatory approach, needs analysis, data-driven decision-making

## INTRODUCTION

In recent years, the University of British Columbia (UBC) has undergone changes in its data governance and started collecting a broader set of student data. As this data became more accessible over the past decade, there has been an increased appetite among faculty members for engaging in data-driven projects and decisions. At the Faculty of Science, the Science Centre for Learning and Teaching, known locally as Skylight (https://skylight.science.ubc.ca/) plays a pivotal role in facilitating systematic planning and ad-hoc learning and teaching data projects, and has become a natural hub for learning and academic data analytics work. This poster describes the evolution of learning analytics work at the Faculty of Science at UBC, highlighting the approach employed by our Centre.

## OUR APPROACH

Through our collaborative work with faculty members and our involvement in teaching and learning initiatives, we have identified numerous opportunities for collecting and utilizing learning and academic analytics-type data. We conducted three sets of needs analyses between 2015 and 2022, adapting our approach as data and tools became available and our understanding of learning analytics-type work evolved. In all iterations, we took a participatory approach (Sarmiento & Wise, 2022) to ensure buy-in, to raise awareness around opportunities related to learning analytics, to gauge and benefit from existing capabilities, to collect information about faculty members and department needs, interests, and priorities, and ultimately to establish a common understanding and agreed-upon needs. We used the information and insights collected from these needs analyses to identify specific objectives, which were then prioritized based on their urgency, feasibility and impact. This was followed by designing analytics tools, running a series of analyses, and generating reports to address the needs. We checked with stakeholders along the way to ensure alignment with their needs and to evaluate the efficacy of the designed tools. Collecting and responding to feedback was integrated as part of this work to ensure efficiency (Rehrey et al, 2020). Figure 1 illustrates this approach.

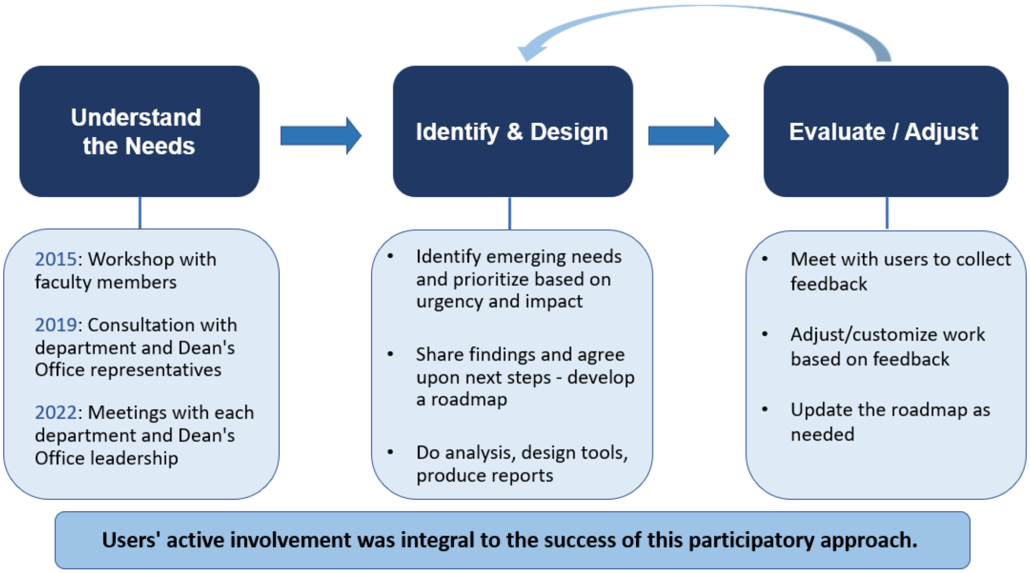



**Figure 1: Participatory approach adopted by UBC Faculty of Science for developing analytics work**

Initial consultations started in 2015 with a workshop that was attended by 35 faculty members from all Science departments. The goal was to demystify learning analytics, introduce the core methods and concepts, and to engage them to get a better sense of their interests in this emerging field. Building on this initial event, we explored ways to gather more information around specific needs and to see if there was an evolution of needs over time. In January 2019, we consulted with our unit's advisory council of faculty consisting of departmental representatives, requesting from them to bring forward their department's needs. Every time we gathered more information, we looked for ways to move the work forward by offering departments data and analyses to keep the momentum and interest going. Around the same time, we engaged with the faculty members in designing and developing a number of learning technologies that mine learning data and provide actionable information about aspects of learning experience. With the addition of a new team member as research analyst and access to data becoming easier, we had enough capacity to run more comprehensive analyses. In the summer of 2022, we conducted our most recent needs analysis with the goal of informing our work and allowing us to better support our departments and programs. As part of this process, we met individually with leadership in all Science departments, programs to discuss their evolving needs. Table 1 summarizes the themes that we identified from our needs analyses.

Informed by the outcomes of this needs analysis, we have developed analysis tools and generated reports on: (a) Gender grade gap analysis in courses, (b) Grade analysis relative to improvements in a large set of courses over a ten-year period, (c) Evaluation of a self-paced academic preparation course for incoming students, (d) Analysis of student course choices relative to first-year course-taking and intended Major, (e) Time-to-degree analysis and (f) Interactive program-specific data reports on course and program profiles.

**Table 1: Emerging themes from the three needs analyses**

| Emerged Themes | 2015 | 2019 | 2022 |
|---|---|---|---|
| Student pathways through programs and courses | ✔ | ✔ | ✔ |
| Impact of teaching initiatives and practices, and new programs | ✔ | ✔ | ✔ |
| Student performance analysis based on background, demographics | ✔ | ✔ | ✔ |
| Historical course data (enrolment, grades, pre-requisite taking, etc.) | | ✔ | ✔ |
| Getting to know students better – more extensive demographics | | ✔ | ✔ |
| Capturing transferable skills such as reasoning and argumentation | ✔ | | |
| Capturing knowledge through concept inventory data to assess student learning | | ✔ | |

## DISCUSSION AND CONCLUSION

The needs analyses conducted between 2015 and 2022 were carried out in different formats and engaged a variety of stakeholders. Our iterative approach helped us ensure our learning analytics work is aligned with the needs of our stakeholders. As we reflect on this journey, it is evident that the participatory approach we took ensured a high level of engagement and buy-in, and this approach remains essential in the continuation of this work. Table 2 provides a high-level summary of commonalities, gains, and lessons learned from engaging in this process.

**Table 2: Notable commonalities, gains and lessons learned from the three needs analyses**

| Commonalities | Gains | Lessons Learned |
|---|---|---|
| High level of engagement and interest in the field | The approach is collaborative, and increases buy-in to the project | Keeping our faculty leadership consistently well-informed |
| Prioritization of academic analytics over learning analytics | Able to highlight the departments' interests to the Dean's Office (and vice versa) | Identifying communication strategies to share the tools |
| Workload and limited capacity at departments | Able to gain a high-level understanding of needs - Knowing what the needs are framed the focus of our future work | Patience! Our interest and ambition were ahead of the institution |

## REFERENCES

Rehrey, G., Molinaro, M., Groth, D., Shepard, L., Bennett, C., Code, W., Reynolds, A., Squires, V., & Ward, D. (2020). Supporting Faculty Adoption of Learning Analytics within the Complex World of Higher Education. In: Ifenthaler, D., Gibson, D. (eds) Adoption of Data Analytics in Higher Education Learning and Teaching. *Advances in Analytics for Learning and Teaching. Springer, Cham*. https://doi.org/10.1007/978-3-030-47392-1_12

Sarmiento, J. P., Wise, A. F. (2022). Participatory and co-design of learning analytics: An initial review of the literature. In LAK22: 12th international learning analytics and knowledge conference, 535-541. https://doi.org/10.1145/3506860.3506910

# The Adoption of Learning Analytics in Blended Learning Environments: An Exploratory Study in Singapore

**Xiao-Feng Kenan Kok, Ching Yee Pua, Shermain Puah, Oran Zane Devilly, Peng Cheng Wang, Eric Chern-Pin Chua**
Singapore Institute of Technology
kenan.kok@singaporetech.edu.sg, chingyee.pua@singaporetech.edu.sg, shermain.puah@singaporetech.edu.sg, oran.devilly@singaporetech.edu.sg, victor.wang@singaporetech.edu.sg, eric.chua@singaporetech.edu.sg

**ABSTRACT**: In the aftermath of COVID-19, blended learning (BL) has emerged as the dominant mode of learning. With our university transitioning to a new campus where BL is expected to seamlessly integrate into its physical design, it is anticipated to become an indispensable component of the student experience. However, there appears to be a paucity of research into students' perceptions of digital support and satisfaction with interactions within BLEs at higher education, especially in the post-COVID-19 era. In this paper, we utilize the learning analytics (LA) cycle as an overarching framework to shape our study's methodology. Our aim was to examine students' perceptions of digital relatedness support (DRS) and satisfaction with learner-technology interaction (SLTI) across three modules ($n$ = 305) at our university. To achieve this objective, we conducted a cross-sectional survey complemented by focus group discussions (FGDs). Over 50% of respondents had positive perceptions of DRS and SLTI. In addition, statistically significant differences were observed between two pairs and one pair of modules for DRS and SLTI, respectively. The FGD data provided some possible reasons for the differences in DRS and SLTI scores across modules. This paper concludes with some recommendations to enhance DRS and SLTI.

**Keywords**: learning analytics, blended learning, digital relatedness support, satisfaction with learner-technology interaction

## 1 INTRODUCTION

Blended learning (BL) is the intentional blending of classroom in-person instruction with digital learning experiences, where student learning is aligned with course objectives (Garrison & Vaughan, 2008). With the commencement of our university's transition to the new Punggol campus in 2024, BL is expected to seamlessly integrate into the student experience. The new campus will feature smaller, modular study halls explicitly designed to facilitate BL, fostering vibrant discussions and collaborative endeavors (SIT, 2022). Despite the normalization of BL as the predominant way of learning post-COVID-19, however, there appears to be limited research into students' perceptions of the BL environment (BLE) (Syska & Pritchard, 2023). Specifically, a paucity of research exists around students' perceptions of digital support (Chui, 2021) and satisfaction with interactions (Gao et al., 2020) in a post-COVID-19 BLE.

The present study uses the learning analytics (LA) cycle (Clow, 2012) as an overarching framework to shape the study's methodology. We aim to examine students' perceptions of digital relatedness support (DRS) and satisfaction with learner-technology interaction (SLTI) across three modules at our university. Table 1 illustrates how we used the LA cycle to craft our research questions (RQs).

**Table 1: Learning analytics cycle, considerations, and research questions.**

| LA Cycle | Considerations | Research Questions |
|---|---|---|
| Learners | Students taking modules that adopt the flipped classroom approach (our institution's default model) to blended learning. | (1) How do learners perceive the current state of blended learning at our university, in terms of DRS and SLTI? |
| Data | Cross-sectional survey complemented with FGDs. | (2) Are there significant differences in students' mean scores of DRS and SLTI across modules? What are some possible reasons for these differences, if any? |
| Metrics / Analytics | Descriptive statistics (e.g., mean, SD), one-way ANOVA, coding of transcripts. | |
| Interventions | Review findings and make recommendations for stakeholders to act upon. | (3) What specific actions can be recommended to enhance, if any, the DRS and SLTI? |

## 2 METHODOLOGY

The method for this exploratory study was a cross-sectional survey design complemented by student focus group discussions (FGD). This study was part of a larger study on blended learning at our institution. Participants were conveniently sampled from three modules ($n$ = 305); Module 1 ($n$ = 74), Module 2 ($n$ = 97), and Module 3 ($n$ = 134). For this study's purposes, we will only report findings from the DRS and SLTI scales in our survey, each containing three items. Cronbach's alpha values were above .70 for DRS and SLTI. The survey data was analyzed using descriptive statistics (i.e., percentage of agreement, mean, SD) and one-way ANOVA to investigate if significant differences existed across modules for DRS and SLTI. We transcribed the FGD recordings verbatim, coded the data, and extracted selected quotes to explain the differences in DRS and SLTI across the three modules. There was a high level of agreement between authors one and two who coded the transcripts.

## 3 RESULTS

The results from RQ1 showed that more than half of the participants, on average, agreed and strongly agreed that the digital tools made them feel more connected to their peers (item 1), instructors (item 2), and helped them relate to the content better (item 3) for DRS. Similar trends were observed for SLTI, where roughly 70% of respondents, on average, enjoyed working in online environments (item 1), felt that such environments made them more productive (item 2), and were very confident in their abilities to navigate in the online environment (item 3).

Regarding RQ2, statistically significant differences were found between Modules 1 and 3 (0.54451, 95% CI [−0.8148, −0.2742]), and Modules 2 and 3 (0.31464, 95% CI [−0.5634, −0.0658]) for DRS mean scores, while SLTI was significantly different between Modules 1 and 3 (0.30234, 95% CI [0.0482, 0.5565]) only. Table 2 illustrates the results of the one-way ANOVA test. The FGD data showed that students taking Module 3 made superficial use of the digital tools (e.g., reading online textbook) and did not engage with them extensively. In contrast, students in Module 2 reported that their instructors used digital tools to interact and engage them, enabling closer connection with peers and instructors. However, students taking Module 1 used digital tools to generate ideas to get started with their projects and showed a strong awareness of the affordances and limitations of digital tools.

The variations in how participants interact with digital tools may elucidate the notable disparities observed in DRS and SLTI among the three modules.

**Table 2: One-way ANOVA results.**

| *Measure* | *Module 1* M (SD) | *Module 2* M (SD) | *Module 3* M (SD) | $F(2, 302)$ | $\omega^2$ |
|---|---|---|---|---|---|
| DRS | 3.96 (.71) | 3.73 (.79) | 3.41 (.84) | 12.027** | .074 |
| SLTI | 4.06 (.68) | 3.82 (.83) | 3.76 (.72) | 4.050* | .026 |

**NOTE:** $^{**}p < .01$, $^{*}p < .05$

## 4 IMPLICATIONS AND CONCLUSION

In response to RQ3, we recommend that the use of digital tools be incorporated into the pedagogical design of the module via developed frameworks (e.g., Väätäjä & Ruokamo, 2021) so that the use of these tools is theoretically grounded. A variety of digital tools could also be used within a module to provide more avenues for engaging with peers and instructors and to meet the diverse digital needs of students. Furthermore, providing opportunities for students to recognize the affordances and limitations of various digital tools could foster greater self-awareness and self-regulation within the BLE. We also suggest that there should be an appropriate balance of face-to-face and online use of digital tools in the BLE. Finally, it would be helpful for instructors to "close the loop effectively" (Clow, 2012, p. 134) with students by explicitly sharing data on DRS and SLTI and showing any changes that have been made in the module based on the interpretation of the data.

## REFERENCES


Chiu, T. K. (2021). Digital support for student engagement in blended learning based on self-determination theory. *Computers in Human Behavior*, *124*, 106909. https://doi.org/10.1016/j.chb.2021.106909

Clow, D. (2012, April). The learning analytics cycle: closing the loop effectively. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 134–138). https://doi.org/10.1145/2330601.2330636

Gao, B. W., Jiang, J., & Tang, Y. (2020). The effect of blended learning platform and engagement on students' satisfaction——the case from the tourism management teaching. *Journal of Hospitality, Leisure, Sport & Tourism Education*, *27*, 100272. https://doi.org/10.1016/j.jhlste.2020.100272

Garrison, D. R., & Vaughan, N. D. (2008). *Blended learning in higher education: Framework, principles, and guidelines*. San Francisco, CA: Jossey-Bass.

Singapore Institute of Technology (SIT). (2023, May 22). *SIT marks Punggol campus topping out milestone with Oracle and Salesforce partnerships*. SIT News. https://www.singaporetech.edu.sg/SITToppingOut

Syska, A., & Pritchard, C. (2023). Blended learning as a site of struggle: A critical realist analysis of students' perceptions of blended learning and its impact on their sense of belonging. *Journal of University Teaching & Learning Practice*, *20*(6), 15. https://doi.org/10.53761/1.20.6.15

Väätäjä, J. O., & Ruokamo, H. (2021). Conceptualizing dimensions and a model for digital pedagogy. *Journal of Pacific Rim Psychology*, *15*. https://doi.org/10.1177/1834490921995395

# Using generative artificial intelligence to scaffold students' metacognitive awareness

**Joni Lämsä[1], Xinyu Li[2], Olli Aksela[1], Matti Kaarlela[1], Mladen Rakovic[2], Dragan Gasevic[2], & Sanna Järvelä[1]**
[1] Learning and Educational Technology Research Lab, University of Oulu, Finland
[2] Centre for Learning Analytics, Monash University, Australia
joni.lamsa@oulu.fi

**ABSTRACT**: If students are not metacognitively aware of the situational demands, they cannot successfully regulate their learning. Metacognitive scaffolds can be used to facilitate this awareness. While metacognitive scaffolds have been personalized based on real-time log traces of self-regulated learning processes, few studies have considered students' metacognitive strategy knowledge in personalizing these scaffolds. Our aim is to study whether and how generative AI can use information on secondary education students' metacognitive strategy knowledge and real-time log traces to personalize metacognitive scaffolds in a multiple-source writing task. We designed prompts for OpenAI's GPT-4 to generate metacognitive scaffolds in three phases of the learning process: at the beginning to facilitate task understanding; in the middle to facilitate monitoring reading; at the end to facilitate monitoring writing. First, we only gave GPT-4 information on the cognitive and metacognitive processes that students (*N* = 20) implemented based on log traces. Second, we supplemented this information with students' metacognitive strategy knowledge. We found that GPT-4 can generate scaffolds to raise metacognitive awareness based on the real-time log traces, and these scaffolds can be further personalized by providing information on the students' metacognitive strategy knowledge.

**Keywords**: artificial intelligence, ChatGPT, metacognition, scaffold, self-regulated learning

## 1 INTRODUCTION

The importance of self-regulated learning (SRL) for learning performance is widely recognized in the AI era (Järvelä et al., 2023). When students successfully regulate their learning, they can strategically apply learning strategies that respond to the situational demands at hand. If students are not metacognitively aware of the situational demands, they cannot successfully regulate their learning. To increase students' metacognitive awareness, metacognitive scaffolds, such as guiding or reflective questions, have been applied to invite cognitive and metacognitive processes (Guo, 2022). A challenge here is that students' metacognitive strategy knowledge varies: Students with low metacognitive strategy knowledge may not benefit from mere awareness of the situational demands for adaptation if they lack the knowledge to control their learning strategies accordingly. At the same time, students with high metacognitive strategy knowledge may only need a discreet remark on the situational demands to take appropriate control (Guo, 2022). Existing studies have designed personalized scaffolds based on real-time log traces of SRL processes (van der Graaf et al., 2023), but few studies have considered students' metacognitive strategy knowledge in personalizing metacognitive scaffolds. Our aim is to study whether and how generative AI can use information on secondary education students' metacognitive strategy knowledge and real-time log traces to personalize metacognitive scaffolds in a multiple-source writing task. Our research question (RQ) is: What is the content of the metacognitive scaffolds when generative AI uses information from 1) students' log traces; 2) the students' metacognitive strategy knowledge and their log traces to generate the scaffolds in a multiple-source writing task?

## 2 METHODS

The participants of the study were 9th-grade students (N = 20). First, students filled out a questionnaire measuring metacognitive strategy knowledge (ISDIMU, Bannert et al., 2021). Second, students were provided with reading material on the topic of human biology (Azevedo et al., 2022), based on which they were given 45 minutes to write a 200–300-word essay on red and white blood cells. The log traces during the essay task resulted from the operations they enacted in a digital learning environment, including keyboard strokes and all navigational activities of the student's interaction with the learning environment. By applying a rule-based AI algorithm (Fan et al., 2022), we detected a set of cognitive and metacognitive processes (e.g., reading, writing, planning, monitoring) based on the log traces. Moreover, the sentences of essays were classified as rehearsing, translating, or assembling based on the log traces.

We have identified three phases that successful students complete during the task (Authors, 2024): at the beginning, students create accurate task understanding; in the middle, students actively monitor their reading; and at the end, students actively monitor their writing. Since secondary education students faced challenges in these three phases (Authors, 2024), we prompted OpenAI's GPT-4 to raise students' metacognitive awareness with scaffolds of a maximum of 50 words in these phases. In the prompt, we explained the goal of each scaffold and how we understand metacognitive awareness. We operationalized relevant behaviors for each phase based on the log traces (a set of cognitive and metacognitive processes for the first and second phases; essay sentence classification for the third phase). First, we included only students' log traces in the prompt when requesting scaffolds. We had 60 scaffolds in total (three for each student, 20 students in the sample). Second, we included students' log traces and their level of metacognitive strategy knowledge in the prompt when requesting scaffolds. To address our RQ, we coded each main and subordinate sentence of the scaffolds as follows: 1. raising awareness on the processes (not) implemented; 2. raising awareness on when and why to apply the processes; 3. providing explicit strategy suggestion; and 4. providing implicit strategy suggestion. We then calculated the mean and standard deviation for the different codes in the different groups (GPT-4 prompted with log traces vs. prompted with log traces and metacognitive strategy knowledge) and compared the means with a Kruskal-Wallis test.

## 3 FINDINGS AND DISCUSSION

The most common category of sentences for the scaffolds based on log traces was raising awareness on when and why to apply the different (meta)cognitive processes (mean = 2.03, sd = 1.11). These types of sentences enhance students' conditional knowledge (when and why to apply specific processes and strategies), which may also benefit those students who have high procedural knowledge (Schuster et al., 2023). The next common categories were implicit strategy suggestions (mean = 1.29, sd = 0.54) and raising awareness on the processes that students have (not) implemented (mean = 1.13, sd = 0.54). The least common category was explicit strategy suggestions (mean = 0.85, sd = 0.87), which aligns with the design of our prompt aimed at raising metacognitive awareness with guiding or reflective questions. The scaffolds based on metacognitive strategy knowledge, plus log traces, were different for students with low and high metacognitive strategy knowledge. Students with low metacognitive strategy knowledge received more explicit strategy suggestions (mean = 1.37, sd = 1.03) and fewer implicit strategy suggestions (mean = 0.43, sd = 0.68) than students with high strategy knowledge (mean = 0.80 and 1.63, sd = 0.80 and 0.76, respectively). These differences were statistically significant ($W$ = 773, $p$ < 0.001 for implicit strategy suggestions; $W$ = 312, $p$ = 0.030 for explicit strategy suggestions). These results show the potential of generative AI to personalize metacognitive scaffolds since considering students' strategic knowledge may increase their strategic adaptations within the task (Guo, 2022). The information on the students' low metacognitive strategy knowledge influenced the content of the scaffolds, making them more explicit and less implicit, compared to the scaffolds based on log traces only. This personalization can decrease the risk of increasing extraneous cognitive load that implicit strategy suggestions may provoke (Guo, 2022). The content of the scaffolds among students with high metacognitive strategy knowledge did not differ regardless of whether the prompt included information on their strategy knowledge or not. Altogether, our findings show that generative AI can be prompted to personalize scaffolds to raise metacognitive awareness based on the real-time log traces, and these scaffolds can be further personalized by providing information on the students' metacognitive strategy knowledge.

## REFERENCES

Azevedo, R., …, & Cerezo, R. (2022). Lessons Learned and Future Directions of MetaTutor: Leveraging Multichannel Data to Scaffold Self-Regulated Learning With an Intelligent Tutoring System. *Frontiers in Psychology, 13*. https://www.frontiersin.org/articles/10.3389/fpsyg.2022.813632

Bannert, M., Arvaneh, B., & Reith, S. (2021). *ISDIMU – Instrument for Assessment of Self-regulated Learning with Digital Media in the Classroom* [Unpublished Manual]. Technical University of Munich, Germany.

Fan, Y., ..., & Gašević, D. (2022). Towards investigating the validity of measurement of self-regulated learning based on trace data. *Metacognition and Learning*. https://doi.org/10.1007/s11409-022-09291-1

Guo, L. (2022). Using metacognitive prompts to enhance self-regulated learning and learning outcomes: A meta-analysis of experimental studies in computer-based learning environments. *Journal of Computer Assisted Learning*, *38*(3), 811–832. https://doi.org/10.1111/jcal.12650

Järvelä, S., Nguyen, A., & Hadwin, A. (2023). Human and artificial intelligence collaboration for socially shared regulation in learning. *British Journal of Educational Technology*, *54*(5), 1057–1076. https://doi.org/10.1111/bjet.13325

Schuster, C., Stebner, F., Geukes, S., Jansen, M., Leutner, D., & Wirth, J. (2023). The effects of direct and indirect training in metacognitive learning strategies on near and far transfer in self-regulated learning. *Learning and Instruction, 83*, 101708. https://doi.org/10.1016/j.learninstruc.2022.101708

van der Graaf, J., …, & Molenaar, I. (2023). How to design and evaluate personalized scaffolds for self-regulated learning. *Metacognition and Learning*, *18*(3), 783–810. https://doi.org/10.1007/s11409-023-09361-y

# A Network Analysis of Collaborative Prototyping from a Knowledge Creation Perspective


| Ayano Ohsaki | Yuanru Tan | Brendan Eagan | David Williamson Shaffer | Zachari Swiecki |
|---|---|---|---|---|
| Shinshu University | University of Wisconsin-Madison | University of Wisconsin-Madison | University of Wisconsin-Madison | Monash University |
| aohsaki@ohsaki-lab.net | yuanru.tan@wisc.edu | beagan@wisc.edu | dws@education.wisc.edu | zach.swiecki@monash.edu |


**ABSTRACT**: This study proposes a combination of temporal socio-semantic network analysis (tSSNA) and ordered network analysis (ONA) to capture knowledge creation (KC) practices during collaborative prototyping. We analyzed the collaborative prototyping activities of three teams—product designers, service designers, and engineers—and found that our novel combination of tSSNA and ONA could distinguish between the processes of these teams over time. Our results inform the design of learning environments for supporting KC practices.

**Keywords**: Knowledge creation, Prototyping, Temporal socio-semantic network analysis, Ordered network analysis, Collaborative problem-solving

## 1 BACKGROUND AND RESEARCH QUESTION

There is a growing interest in knowledge creation (KC) practices in science, technology, engineering, and mathematics (STEM) education. KC practices address ill-formed problems, necessitating environments that encourage learners to represent and refine ideas through design-mode thinking based on high-level epistemic agency, such as choosing important problems (Chen & Zhang, 2016). However, during such practices, learners often avoid presenting incomplete ideas. We argue that a deeper understanding of what is happening and how it is happening during prototyping—a well-known effective practice in engineering and design that allows for imperfection and awkwardness (IDEO U, n.d.)—can contribute to the design of better learning environments for KC.

Building on prior research, our study introduces a novel combination of temporal socio-semantic network analysis (tSSNA) (Ohsaki & Oshima, 2021) and ordered network analysis (ONA) (Tan et al., 2023) to capture characteristics of the collaborative prototyping processes of three teams: product designers, service designers, and engineers. Our study seeks to address the research question: What are the differences in the prototyping of engineers, product designers, and service designers from a KC perspective, and how do these differences change over time?

## 2 METHODS

We collected data from three teams tasked with prototyping a new wallet in a 30-minute session based on user needs. The user, played by the session facilitator, was a woman in her thirties who aimed to adopt a minimalist and healthy lifestyle. The conversations in the team activity were recorded and transcribed. The teams consisted of three professional graduate school students in software engineering (Engineering), three undergraduate students in product design (Product Design), and three professional service designers (Service Design). The data included 1,345 turns of talk with an average per team of 448 turns (SD = 107).

To address our research question, we first used tSSNA to capture idea improvement. tSSNA extends socio-semantic network analysis (SSNA) (Ohsaki & Oshima, 2021) by incorporating timestamps, moving windows, and network lifetime for real-time events to capture more nuanced temporal variations in collaboration compared to previous methods. The method uses key phrases as nodes to create networks and total degree centrality (TDC) to quantify how ideas change by representing network structure and restructuring (Oshima et al., 2012). The analysis segmented team discussions into three phases based on TDC score decay (Ohsaki & Oshima, 2023), resulting in distinct idea improvement phases for each team. Second, the study used ONA (Tan et al., 2023) to analyze design action patterns via directed network graphs. Two-dimensional vector representations of these graphs—ONA scores—of each team by phase were visualized. We statistically compared the differences between scores on each dimension using fixed-effects regression analysis. Prior to ONA, we coded each utterance for the codes: User (referencing the user), Vision (referencing the designers' plan for the product), Prototyping (referencing the prototype), Function (referencing the product's functions), and Aesthetics (referencing visual/exterior/fitting of the design), based on relevant design theories (Ohsaki & Oshima, 2023).

# 3 RESULTS AND FUTURE RESEARCH

In Figure 1 (left), the X-axis represents time, and the Y-axis represents TDC. The graph illustrates continuous fluctuations in TDC values for all teams throughout the study, indicating consistent idea evolution. The teams initially displayed high scores and frequent oscillations, followed by a temporary decrease and subsequent recovery to high variability and scores. The phases spanned from start to 10 minutes, 10 to 27 minutes, and 27 minutes to the end of the sessions.

The ONA results show each team by phase in the low-dimensional space in terms of their ONA scores. We have interpreted the dimensions of the space according to the connections between codes that distinguish the teams the most (bold terms in Figure 1 (right)). The X-axis distinguishes between teams who made stronger connections to User versus Prototyping; The Y-axis distinguishes between teams who made stronger connections to Aesthetics versus Vision and Functions. Overall, the ONA graph revealed a general trend of teams moving from left to lower-middle to upper-right across the phases, but variations in trajectories indicated unique team dynamics. Regression analyses confirmed several teams differed significantly ($p < 0.05$) between phases and teams in each phase. Regarding phases, Engineering Team differed between Phases 1 and 2 (X), Phases 1 and 3 (X and Y), and Phases 2 and 3 (Y); Product Design Team differed between Phases 1 and 2 (X), Phases 1 and 3 (X and Y), and Phases 2 and 3 (Y); Service Design Team differed between Phases 1 and 2 (Y), Phases 1 and 3 (X and Y), and Phases 2 and 3 (X). Regarding teams in each phase, Service Design Team differed significantly from Engineering Team in Phase 1 (X and Y), Phase 2 (X), and Phase 3 (Y); They differed significantly from the Product Design Team in Phase 2 (X and Y) and Phase 3 (Y).

In this paper, we analyzed the collaborative prototyping processes of three teams using a novel combination of tSSNA and ONA from the KC perspectives of the epistemic agency and design mode thinking including how participants improve ideas, what design action they choose, and how they choose design actions. Our method could distinguish the three phases in each team statistically. The results suggest that our approach can effectively model iterative design activities like KC even for small sample sizes. The analysis focused on idea improvement and design actions based on design theories (Ohsaki & Oshima, 2023), but future work will explore the relationships between designers' shared epistemic agency—an essential aspect of KC theory (Scardamalia, 2002)—and design actions.

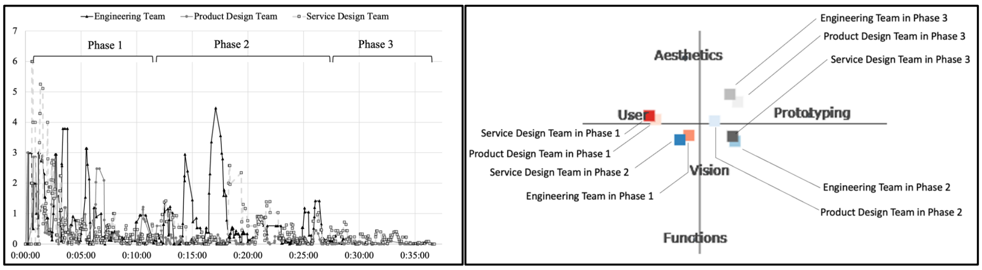



**Figure 1: The results of tSSNA and ONA (left: the transitions of total degree centrality in tSSNA, right: ONA scores for teams by phase.)**

## ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Numbers 18K13238, 22H01043 and 23K11357. This work was funded in part by the NSF (DRL-2100320, DRL-2201723, DRL-2225240), the Wisconsin Alumni Research Foundation, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison. The opinions, findings, and conclusions do not reflect the views of the funding agencies, cooperating institutions, or other individuals.

## REFERENCES

Chen, B., & Zhang, J. (2016). Analytics for knowledge creation: Towards epistemic agency and design-mode thinking. *Journal of Learning Analytics*, 3(2), 139-163.

IDEO U (n.d.). Prototyping elmo's monster maker app. Retrieved from https://www.ideou.com/blogs/inspiration/prototyping-elmo-s-monster-maker-app

Ohsaki, A., & Oshima, J. (2021). Socio-semantic network analysis of knowledge-creation discourse on a real-time scale. In: Ruis A. R., & Lee S. B. (Eds.) *Advances in Quantitative Ethnography* (pp. 170–184). Springer, Cham.

Ohsaki, A., & Oshima, J. (2023). Analyzing the co-design process by engineers and product designers from perspectives of knowledge building. In: Damşa, C., & Barany, A. (Eds.) *Advances in Quantitative Ethnography* (pp. 388–401). Springer, Cham.

Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: A social network analysis application for knowledge building discourse. *Educational Technology Research and Development*, 60, 903–921.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. Liberal education in a knowledge society, 97, 67–98

Tan, Y., Ruis, A. R., Marquart, C., Cai, Z., Knowles, M. A., & Shaffer, D. W. (2023). Ordered network analysis. In: Damşa, C., & Barany, A. (eds) *Advances in Quantitative Ethnography* (pp. 101-116). Springer, Cham.

# Designing a Dashboard to Enhance Reflection of Microteaching in Multi-user Virtual Reality application for Pre-Service Teachers

**Martin Hlosta[1*], Ivan Moser[1], Per Bergamin[1], Umesh Ramnarain[2], Christo van der Westhuizen[2], Mafor Penn[2], Noluthando Mdlalose[2], Koketso Pila[2], Ogegbo Ayodele**

[1] Swiss Distance University of Applied Sciences (FFHS), Switzerland
[2] University of Johannesburg, South Africa
[*] martin.hlosta@ffhs.ch

**ABSTRACT**: Despite a growing number of cases exploring the potential of Immersive Virtual Reality in education, contributions trying to utilise rich multisensory data coming from these devices are lacking. This paper reports on the development of a data pipeline from a multiuser VR application designed for pre-service teachers to practise their skills using microteaching. In parallel, it presents a learning analytics dashboard and its co-design with the educators and follows the needs of the underlying pedagogy model of microteaching. The essential part of the exercise is the reflection that follows the microteaching, and the dashboard focuses on enhancing the feedback of this reflection. Both the VR application and the dashboard itself are now deployed and being piloted with the first cohort of students (pre-service teachers) in South Africa (N=30). The main aim of this practitioner-track paper is to fill the gap in reporting on the development of a dashboard for the educational VR application, which despite its promise is non-existent in reported research or exists only as a concept. We do so by sharing the concept of the pipeline and the decisions made during the dashboard design, which could help future practitioners bypass the initial knowledge gaps.

**Keywords**: Virtual Reality, Multimodal Learning Analytics, Pre-Service Teachers, Dashboards.

## 1 BACKGROUND

Use cases of piloting Immersive Virtual Reality (VR) for education are growing, both in formal education and professional development, including evidence of higher interest and self-efficacy in classroom management (Huang et al., 2023). VR devices allow collecting rich sources of multisensory data (e.g. positional data, eye-tracking), opening new possibilities to provide feedback via Learning Analytics (LA). Despite the growing interest in VR and its convergence with LA, only a few examples such as (Heinemann et al., 2023) have been published. This may be related to the challenges of VR - the sensory data is more complex and poses new data engineering challenges to obtain good quality data. Also, the richness of the sensory data poses new privacy and security challenges (Kukulska-Hulme et al, 2023). LA Dashboards (LADs) are a typical way of communicating the analytics to stakeholders. While some have shown promising effects, e.g. increased student retention (Herodotou et al, 2023), they are often criticised for not involving users in the co-design (Nazaretsky et al., 2022), or not being grounded in learning theories (Dourado et al., 2021). Despite the lack of research on LAD for VR, several papers focus on the integration and presenting data from various multimodal sensors. They mostly focus on combining the extraction of information from video and additional physical sensors.

## 2 LAD DESIGN PROCESS

The motivation for designing the VR application with LAD came from discussions with teaching practitioners about the challenges of pre-service education in STEM subjects. In STEM, a teacher is often challenged to explain a complex phenomenon, and 3D visual objects can be useful to complement the verbal explanation and facilitate understanding. At the same time, the use of VR avoids the need to purchase of expensive lab equipment to demonstrate these concepts. A noteworthy aspect is that the target group is in South Africa, with scarce resources, but with government support to see technology to overcome these challenges. **Focusing on pre-service teachers (PSTs)** may allow them to familiarise themselves with the use of technology in their future teaching. The application focuses on bringing the concept of micro-teaching, i.e. short practice lessons usually recorded on video, into VR. The key element of micro-teaching is the subsequent reflection (Amobi, & Irwin, 2009), aiming to improve the practice of pre-service teachers. LA was identified as a means of enhancing this feedback by collecting and providing objective data via a LAD.

In response to the LADs without involving stakeholders and pedagogical grounding, a four-day workshop was organised to design the concept of the VR application and conceptualise the LAD. The team of researchers and practitioners identified the **instructional design aligned with the VR application** and STEM based on inquiry-based learning using the 5E model (Duran & Duran, 2004). In the VR application, PSTs deliver a ~15-minute lesson to their peers, with the educator being an invisible observer. The lesson consists of a phenomenon introduction, working in groups, and a final discussion. Web interface minimises the time students need to spend in VR, and reduces the amount of discomfort, which some people experience as a result of prolonged time in VR. Previous research identified that students and teachers prefer actionable LADs with recommendations on how to improve. But since research of teaching and analytics in VR is fairly unexplored, the first pilot focuses on a **descriptive LAD.** We also collect educators' feedback via newly designed rubrics, paving a way to identify measurable performance indicators for diagnostic analytics and actionable feedback.

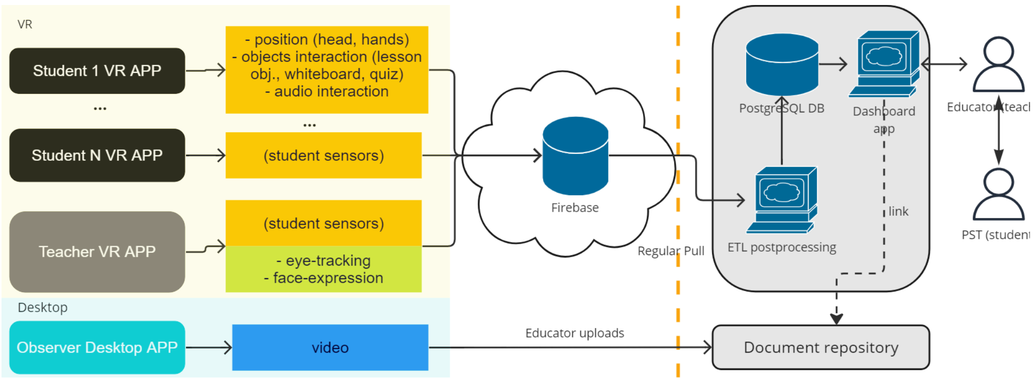



**Figure 1: Data pipeline of the VR application to the LA Dashboard**

## 3 DATA PIPELINE

The VR application supports two different headset devices - the PSTs in the student role wear Meta Quest 2, while the teacher receives a Meta Quest Pro. Both devices record the positions and rotations of the headset and hand controllers, audio, and any interactions with the 3D objects. The Meta Quest Pro also provides eye and face tracking data. Eye tracking allows studying where the teacher directs their attention while delivering the lesson. Face tracking enables the application to convey cues of the teachers' emotional state to the students, aiming to increase the social communication experience. To enhance accessibility, we designed an **automated pipeline** for using the sensors and handling the recorded data, depicted in Figure 1. Each device collects pseudonymised sensory data and sends them securely to Google Firebase DB (selected for easy integration with the VR application). The data is sent in small batches, allowing us to collect data at 50 millisecond resolution, important for the eye-tracking sensors. Lesson data is transferred to our internal server for pre-processing. This consists of merging and aggregation on different time-granularity for presentation in various timeline graphs. The visualisations are made accessible in the LAD within 15 minutes of lesson completion.

## 4 REFLECTIONS

The biggest challenges involved the efficient data delivery to the teacher with minimum effort on their part. The project is now piloted at University of Johannesburg in South Africa with 30 PSTs, with obtained ethical approval. Each student should teach 1 lesson, and act as a peer student for several others. The aim is to investigate both the efficiency of using VR for teaching and how the LAD can enhance the feedback in these lessons. The first piloting round gathers data on which measurable elements in the VR micro-teaching are associated with higher rubric performance. At the same time, qualitative feedback should enhance the design for the next two rounds of piloting.

## REFERENCES


Amobi, F., & Irwin, L. (2009). Implementing on-campus microteaching to elicit preservice teachers' reflection on teaching actions: Fresh perspective on an established practice. Journal of the Scholarship of Teaching and Learning, 27-34.

Duran, L. B., & Duran, E. (2004). The 5E instructional model: A learning cycle approach for inquiry-based science teaching. *Science Education Review*, *3*(2), 49-58.

Dourado, R. A., Rodrigues, R. L., Ferreira, N., Mello, R. F., Gomes, A. S., & Verbert, K. (2021). A teacher-facing learning analytics dashboard for process-oriented feedback in online learning. In LAK21: 11th International Learning Analytics and Knowledge Conference (pp. 482-489).

Heinemann, B., Görzen, S., & Schroeder, U. (2023). Teaching the basics of computer graphics in virtual reality. *Computers & Graphics, 112*, 1-12. https://doi.org/10.1016/j.cag.2023.03.001

Herodotou, C., Maguire, C., Hlosta, M., & Mulholland, P. (2023). Predictive Learning Analytics and University Teachers: Usage and perceptions three years post implementation. In LAK23: 13th International Learning Analytics and Knowledge Conference (pp. 68-78).

Huang, Y., Richter, E., Kleickmann, T., & Richter, D. (2023). Comparing video and virtual reality as tools for fostering interest and self-efficacy in classroom management: Results of a pre-registered experiment. British Journal of Educational Technology, 54(2), 467-488.

Kukulska-Hulme, A., Bossu, C., Charitonos, K., Coughlan, T., Deacon, A., Deane, N., Ferguson, R., …, & Whitelock, D. (2023). Innovating Pedagogy 2023: Open University Innovation Report 11. Milton Keynes: The Open University.

Nazaretsky, T., Bar, C., Walter, M., & Alexandron, G. (2022). Empowering teachers with AI: Co-designing a learning analytics tool for personalized instruction in the science classroom. In LAK22: 12th International Learning Analytics and Knowledge Conference (pp. 1-12).

# Learning Analytics for Ubiquitous Learning: Linking In-class and Out-of-class Contexts

**Manabu Ishihara, Yuko Toyokawa, Izumi Horikoshi, Hiroaki Ogata**
Kyoto University
ishihara.manabu.46z@st.kyoto-u.ac.jp, yflamenca18@gmail.com,
horikoshi.izumi.7f@kyoto-u.ac.jp, hiroaki.ogata@gmail.com

**ABSTRACT**: School educators occasionally interweave classroom and field activities, yet it tends to be difficult to form meaningful links between the different contexts. To overcome that, combining Ubiquitous Learning technologies and Learning Analytics seems beneficial. Current studies, however, have not yet articulated the ways of collecting and leveraging the learning logs across the contexts. This study attempted to grab insights from the learning logs based on a mock English class consisting of two in-class and one out-of-class activities. Our cross-context data analysis relying on network analysis found insights at the individual level.

**Keywords**: Ubiquitous Learning, Cross-context data analysis, In-class and out-of-class learning

## 1 INTRODUCTION

School educators occasionally implement field activities in addition to the classroom learning, for instance, field trips. To exploit the advantages, it is crucial to effectively bridge in- and out-of-school learning (Eshach, 2007). Despite the use of field notes and reflection, it would still be hard to clarify the relevance between the two learning contexts. The combination of Ubiquitous Learning technologies and Learning Analytics (LA) that aims at fostering learning in any situations could address the issue. Although a few works achieved visualization (Mouri et al., 2018), scarce study has tackled the analysis of learning logs across classroom and field activities to ensure students' meaning making. This study thus addresses the issue with the following research question: Is it possible to grab insights regarding one's learning process across in-class and out-of-class contexts?

## 2 METHOD

### 2.1 Data Collection

We employed an LA platform including an e-book reader called BookRoll and a mobile web application called SCROLL (Flanagan & Ogata, 2018) to accumulate learning logs from in-class and out-of-class contexts. SCROLL was developed especially for real-world language learning. It allows learners to find and save foreign words as pairs with photos of the scenes where the words were found.

We then conducted a mock English class consisting of a pre-classroom activity (in-class context), a field activity (out-of-class context), and a post-classroom activity (in-class context). Six graduate students and one faculty member participated. None of them were native English speakers. All participants knew that the class was designed for a cross-context data analysis, but they did not know the research question nor the protocol of data analysis. From the two in-class contexts, we collected the text logs derived from participants' use of the marker and the memo functions on BookRoll. The vocabulary in the in-class material, shown on BookRoll, was also collected. As for the out-of-class context, we collected the logs on SCROLL, namely, pairs of a word and a photo (Figure 1 a).

The accumulated learning logs were anonymized, tokenized and normalized by natural language processing (NLP) techniques, and then structured suitable for network analysis (Figure 1 b). We then built each participant's network to answer the research question at the individual level (Figure 1 c).

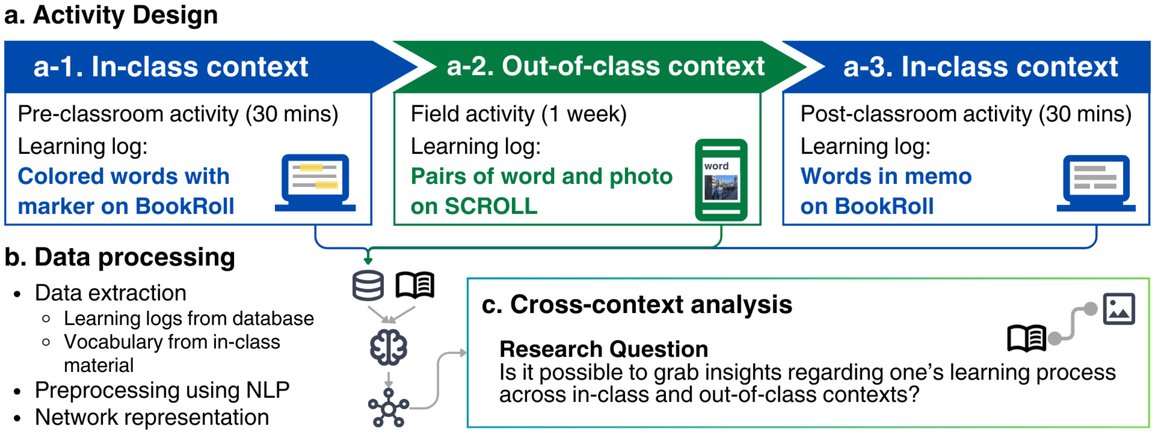



**Figure 1: Method overview**

## 2.2 Activity Design

In the first in-class context, the participants met the learning goal: introduce a place in English. Then they read example English sentences in the material on BookRoll. After that they used the marker function to identify unfamiliar words (Figure 1 a-1) and planned a visit to a location where they could pair the words with scenery photos. Additionally, they were instructed to find further pairs using unplanned words in parallel with the main task. This additional task intended to capture participants' improvisational interests during their field activity. The out-of-class context took place for one week following the previous context. The participants visited their planned locations and saved pairs of a word and a photo using SCROLL on their smartphone (Figure 1 a-2). The use of other tools such as search engine was not suggested but not limited. In the second in-class context, the participants used the memo function on BookRoll for a reflection and a composition to introduce the visited place (Figure 1 a-3). In the end, each participant gave an oral presentation based on the memo.

# 3 RESULTS AND DISCUSSION

Although this study involved a small number of participants, the research question is still worth discussing as it concerns the feasibility of cross-context data analysis. Our networks depicted the links between participant, word, photo, and in-class material nodes in each context (Figure 2). Analyzing the appearance of nodes and edges would suggest how one's use of words was widened and deepened. Also, that would be clues to assess if one's footprints aligned with the activity design.

In the case of participant S02, for example, the network in the first in-class context depicted that the participant identified thirteen unfamiliar words (Figure 2 left). Then, five of them were paired with photos during the out-of-class context (Figure2 middle). These five words could represent particular focus of the participant. Otherwise, it could represent the discontinuance of the field activity. The words used in the field activity are a subset of the planned words in the first in-class context. It means there was no pair for the additional task, which expected unplanned word usage. In short, participant S02 might holt the main task after achieving the fifth pair and did not tackle the additional task.

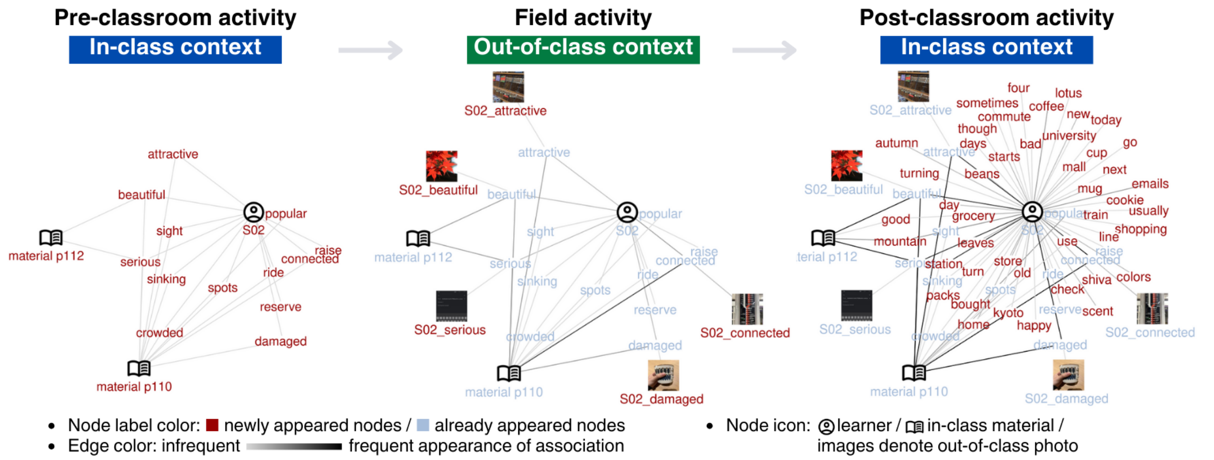



**Figure 2: Example individual network (case of participant S02)**

As for the second in-class context, there were 45 newly appeared nodes derived from the composition activity (Red nodes in Figure2 right). The operation of vocabulary would be insightful in the sense of estimating one's writing proficiency. Rather than just an individual analysis, comparative analysis among participants would also be intriguing. For instance, co-occurrence of words would highlight group awareness. On the other hand, the insights from density of edges are valuable, too. The darkest edges imply that the associations, including the five words used in the field activity, repeatedly appeared in all contexts (Black edges in Figure 2 right). They can be seen as the hubs across contexts, a key for meaning making. Such clues could be part of modeling for intelligent assistance that accordingly reminds relevant field experiences during a classroom activity and vice versa.

## 4 FUTURE WORK

Our approach revealed at which point one's learning process got sparked, deepened, or stuck across classroom and field activities. To go further, we need to examine the applicability of the approach to other learning objectives and group-level analysis. It is also vital to identify insightful logs besides text format. Lastly, ethical concerns along with visual and location data should be discussed.

## ACKNOWLEDGEMENT

We appreciate the generous advice from Dr. Noriko Uosaki. This work was supported by CSTI SIP Program Grant Number JPJ012347.

## REFERENCES

Eshach, H. (2007). Bridging In-school and Out-of-school Learning: Formal, Non-Formal, and Informal Education. *Journal of Science Education and Technology*, *16*(2), 171–190.

Flanagan, B., & Ogata, H. (2018, July). *A Learning Analytics Platform Approach to Seamless Learning*. Paper presented at the 7th International Congress on Advanced Applied Informatics (IIAI-AAI).

Mouri, K., Uosaki, N., & Ogata, H. (2018). Learning Analytics for Supporting Seamless Language Learning using E-book with Ubiquitous Learning System. *Journal of Educational Technology & Society*, *21*(2), 150–163.

# Psychometric Modeling of Speed and Accuracy: Analysis of the National Assessment of Digital Literacy Data in South Korea

**Hyelim Noh[1], Hae Won Cho[1], Nara Han[2], & Hyun Sook Yi[1*]**
[1]Konkuk University, South Korea
[2]Korea Education and Research Information Service, South Korea
ahdclal5@naver.com, johyew1@gmail.com, nara@keris.or.kr, hyunsyi@konkuk.ac.kr[*]

**ABSTRACT**: This study investigates the relationship between response time and accuracy among South Korean students using the National Assessment of Digital Literacy, a digital literacy test designed for a nationally representative sample of students from grades 4 through 9. The hierarchical model proposed by van der Linden (2007) was employed to model the relationship between response time and accuracy. Female students exhibited slightly higher scores than male students, while the mean response time per item remained comparable across gender groups in all subdomains. Average scores increased and response time decreased as the grade level advanced. Students using PCs, as opposed to mobile devices, exhibited higher accuracy and faster response speed. Male students exhibited stronger latent correlations between speed and accuracy.

**Keywords**: Digital literacy, response time, hierarchical model for speed-accuracy relationship

## 1 INTRODUCTION

In our technology-driven society, it is imperative that individuals of all ages possess the essential skills to comprehend, utilize, and critically assess digital information. Recognizing the importance, South Korea has developed and implemented the National Assessment of Digital Literacy (NADL), a digital literacy test designed for a nationally representative sample of students from grades 4 through 9. The NADL is a large-scale web-based test consisting of 26 scenario-based performance questions that assess five subdomains of digital literacy: (A) Digital tools, (B) Digital information & data, (C) Digital communication & collaboration, (D) Production of digital resources, and (E) Digital safety and health (Kim et al., 2023). Students' responses to each item and their interactions with the assessment system are recorded in student logs, with timestamps for specific actions.

Response time serves as a valuable indicator for checking the credibility of student responses (Yamamoto & Lennon, 2018) and for deducing the level of engagement exhibited by students during the test (Goldhammer et al., 2016). Previous studies examining the relationship between speed and accuracy in performing cognitive tasks identified a phenomenon known as the speed-accuracy tradeoff, indicating a tendency for accuracy to decline as speed increases (i.e. response time decreases). However, recent studies propose that this relationship is contingent on the specific characteristics of assessments as well as the population of examinees (Shin, 2021). In PISA and ICILS, two globally recognized international assessment programs, South Korean students exhibited positive relationships between accuracy and speed, showcasing high average performance in the domains being measured and exceptionally rapid average response times compared to participating countries (OECD, 2020). These relationships were more pronounced among male students than female students.

Research has been sparse in modeling and measuring the intricate relationships between student performance and response time, especially in low-stakes testing programs such as the NADL. Consequently, this study aims to investigate the relationship between response time and accuracy among South Korean students using 2023 NADL data, with a specific focus on gender differences.

## 2 METHOD

We analyzed the relationship between response time and the accuracy of the response among grades 7 through 9 who participated in the 2023 NADL (17,120 students). The hierarchical model proposed by van der Linden (2007) was employed to model the relationship between response time and accuracy. This multidimensional item response theory model consists of measurement models of two latent constructs (speed and accuracy), connected through a higher-level correlation. The latent 'accuracy' is measured by scores on each item within the domain. The latent 'speed' is quantified by the time invested in each test item using a log-normal model, and the parameters of time intensity and time discrimination are subsequently estimated. The relationship between speed and accuracy was compared between gender, grade levels, and digital devices that an individual student used for the test. The analysis was conducted with each subdomain of the NADL, using Mplus 8.4.

## 3 RESULTS

Figure 1 illustrates the mean scores and average response time per each item of the NADL, displayed separately by gender, grade levels, and device. Female students exhibited slightly higher scores than male students, while the mean response time remained comparable across gender groups in all subdomains. As the students move to higher grade levels, mean scores increased and response time decreased. Students who using PCs, as opposed to mobile devices, exhibited higher accuracy and faster response speed.

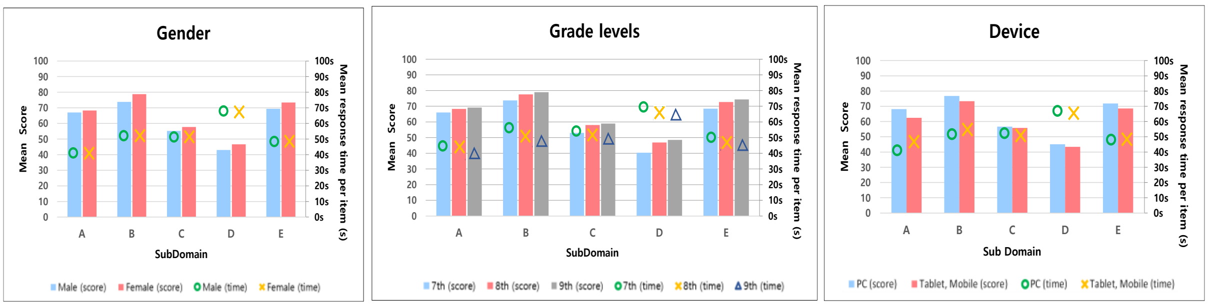



**Figure 1: Mean scores and item response times of the NADL across gender and grade levels**

The leftmost two panels of Figure 2 illustrate the standardized regression weights of device and grade levels on the accuracy and speed. Given that the model-data fit was acceptable for the three subdomains, (A) Digital tools, (B) Digital information & data, and (E) Digital safety & health, further analyses were conducted for these subdomains. Results show that students using PCs demonstrated higher accuracy and the faster response speed for subdomain (A), while students using mobile devices exhibited higher accuracy and slower response speeds for subdomain (B) and (E). As grade levels increased, average scores improved, and response times shortened for the three subdomains.

The final panel of Figure 2 presents the accuracy-speed latent correlations, adjusting for the effects of both device type and grade levels. A positive correlation indicates that students with higher accuracy tend to take longer to respond, as the response speed should be interpreted as slowness in the model. In examining the relationship between accuracy and speed while accounting for grade and device type, subdomains (A), (B), and (E) displayed correlations above 0.5 across all student groups. An analysis segregated by gender revealed that both genders generally took longer to respond when they were more accurate. However, male students exhibited stronger latent correlations compared to female students.

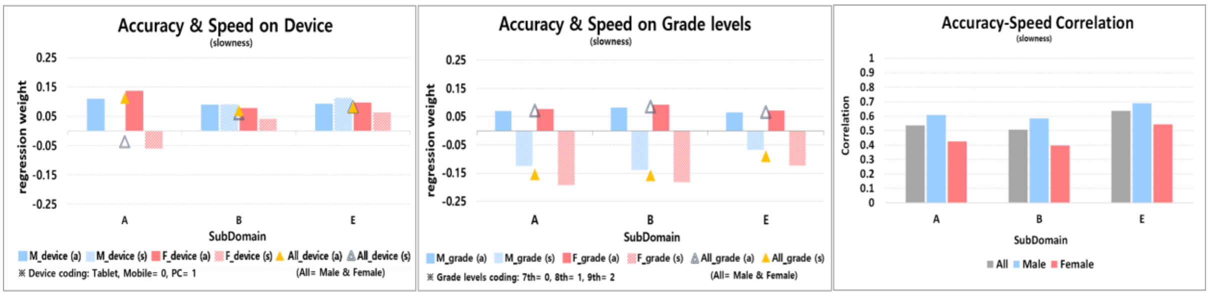



**Figure 2: Regression weights of device & grade levels on accuracy & speed and accuracy-speed latent correlations**

## 4 LIMITATIONS AND FUTURE WORK

The hierarchical model did not demonstrate an acceptable model-data fit for two subdomains of the assessment under consideration in the study. This suggests the necessity of exploring alternative psychometric modeling approaches that better capture the nature of the speed-accuracy relationship.

## ACKNOWLEDGEMENT

This work was supported by Institute of Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No. RS-2023-00229780). The data used in this study were provided by the Korea Education and Research Information Service.

## REFERENCES

Goldhammer, F. Martens, T., Christoph, G., & Lüdtke, O. (2016). Test-taking engagement in PIAAC, *OECD Education Working Papers 133*. Paris, France: OECD Publishing. doi:10.1787/5jlzfl6fhxs2-en

Kim, S., Yi, H., Kim, H., Park, J., Choi, M., Han, N. (2023). *Improvement and Level Measurement of Performance-Based Assessment Tools for Digital Literacy in Elementary and Middle School Students in 2023*(Report No. KR 2023-13). Korea Education and Research Information Service.

OECD (2020). *PISA 2018 Technical Report*. OECD Publishing. http://www.oecd.org/pisa/data/pisa2018technicalreport/

Shin, H. J. (2021). Psychometric modeling of speed and accuracy: Analysis of PISA 2015 data from Korea and the United States. *Journal of Educational Evaluation, 34*(3), 587-614.

van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika, 72*(3), 287-308.

Yamamoto, K., Shin, H., & Khorramdel, L. (2019). Introduction of multistage adaptive testing design in PISA 2018. *OECD Education Working Papers No. 209*, OECD Publishing, Paris. https://dx.doi.org/10.1787/b9435d4b-en

# How Social Comparison Can (Not) Increase Student Motivation

**Author(s): Aditya Joshi, Sergey Sosnovsky**
Utrecht University, The Netherlands
{a.joshi, s.a.sosnovsky}@uu.nl

**ABSTRACT**: Learning Analytics Dashboards (LADs) can inform students about their progress compared to other students. Social comparison (SC) is known to influence students' motivation and learning outcomes. Variations in the intensity and direction of this influence can be attributed to individual differences among students. We conducted an observational study to understand how a SC-enabled LAD using a more modest social norm can influence students' motivation. We found that their Achievement Goal Orientation affects their behavior when they are exposed to such SC. Performance-oriented students are significantly influenced by SC and become susceptible to demotivation in the investigated context. These findings have important implications for the design of LADs employing SC.

**Keywords**: Learning Analytics Dashboard, Social Comparison, Motivation

## 1 INTRODUCTION AND BACKGROUND

Social Comparison (SC) is a psychological phenomenon explaining human tendency to evaluate themselves by comparing with others (Festinger, 1954). Among other effects, SC has been found to provide an important source of motivation. Learning Analytics Dashboards (LADs) often implement SC as a *one-size-fits-all design* (Teasley, 2017) providing a similar interface to all students. Although SC has been shown to improve students' motivation, some students may also lose motivation by feeling either incompetent when faced with results of better-performing peers or accomplished too early when finding themselves performing better than others. A review of LADs (Jivet, Scheffel, Drachsler, & Specht, 2017) underlines that very few interfaces rationalize the use of SC and identify the need to further investigate students' behavior in the presence or absence of SC. At the same time, research shows that people may differ fundamentally in terms of how they engage with SC information. We compare upwards (to others performing better than us) or downwards (to others performing poorer than us); we compare for self-appraisal, enhancement, or improvement (Wood, 1989), etc. There is also a growing body of literature investigating individual differences of students' interaction with and perception of SC-enabled interfaces (Joshi, Molenkamp, & Sosnovsky, 2023) (Akhuseyinoglu, Milicevic, & Brusilovsky, 2022) (Sosnovsky, Fang, Vries, & Luehof, 2020). The impact of SC on motivation can be driven by various factors. In this work, we focus on Achievement Goal Orientation (AGO) (Elliot, 2001) as a potential framework helping understand how SC may motivate or demotivate students. Achievement Goals are categorized as Mastery Goals that focus on learning and improvement, or Performance Goals, that emphasize achieving a good grade and outperforming others. The aim of this work is to study the effect of SC on students' interaction with progress indicators in an LA Dashboard, and how their individual differences modulate this effect. This is addressed by comparing students' engagement with non-mandatory learning material depending on the level of SC information they are shown, and their achievement goal orientation. The impact of SC on motivation can be driven by various factors. In this work, we focus on Achievement Goal Orientation (AGO) (Elliot, 2001) as a potential framework helping understand how SC may (de-) motivate students. Achievement Goals are categorized as Mastery Goals that focus on learning and improvement, and Performance Goals that emphasize receiving a good grade and outperforming others. This work studies the effect of SC on students' interaction with progress indicators modulated by their individual differences. This is addressed by comparing students' engagement with non-mandatory learning material depending on the level of SC information they observe, and their AGO.

## 2 STUDY

We have developed a web-based application called *StudyLens* that supports students' interaction with online learning material. Its interface provides students with an overview of their progress on various topics of the course. Each *Topic Tile* shows an individual *progress bar* indicating how much learning material a student has completed (Figure 1a). The system can augment a tile with a social progress bar displaying the average progress of the class on that topic (Figure 1b). By clicking the tile, a student expands it to reveal corresponding learning activities (Figure 1c). Once an activity is completed (correctly), its icon is marked accordingly and the student's progress for the topic grows.

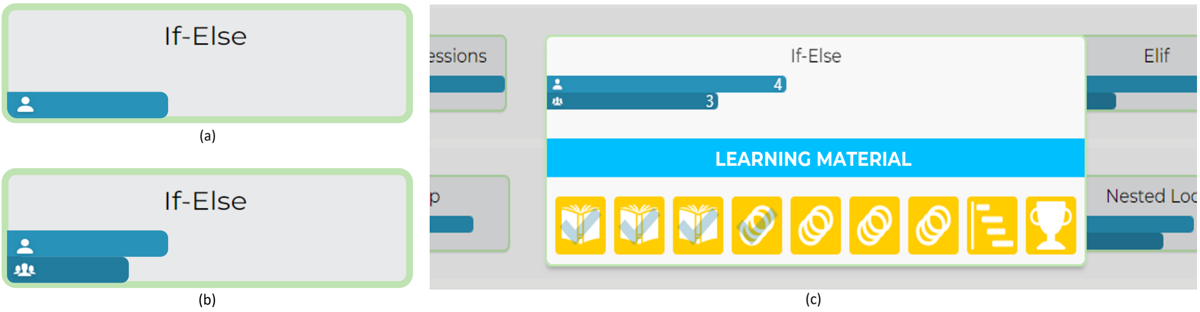



Figure 1: Topic Tiles for a topic in (a) NO SC and (b) SC conditions. (c) Expanded Topic Tile with SC

The experiment was set up in an introductory Python course taught in February 2023 at Utrecht University, Netherlands. The system was introduced to the students as a supplementary tool to practice programming outside of teaching hours. The integrated external learning activities included animated examples, programming challenges, Parson's problems, and instructional readings (Manzoor, Akhuseyinoglu, Shaffer, & Brusilovsky, 2019). All students' interaction with *StudyLens* was processed to infer several parameters characterizing the quantity and quality of work students have done in the system: the number of Activities Started, Activities Revisited, Number of Sessions, and number of Days logged in. The students were randomly assigned to one of the three Conditions – No SC (never see social progress indicators), SC on Demand (see the social progress indicator only when the topic tile is unfolded), and SC (always see social progress indicators for all topics). Out of 165 students enrolled in the course, 117 accessed the system more than once, averaging 150 activities. Only 57 students completed the AGO questionnaire at the beginning of the course. Using the system was not compulsory. Hence, only about half of registered students actively used it. Nevertheless, we decided to compute the average progress using all registered students. Due to this, the average progress of the class across course topics was quite low, and the SC progress indicators consistently displayed a fairly modest social norm. As a result, the overall effect of SC on motivation was not positive. In fact, students who had access to SC were able to "outperform" the social norm very quickly. Consequently, students who observed the social progress indicators *and used them to decide which topics to focus on and which topics to ignore*, were not motivated to work on many learning activities per topic; completing a few was enough to achieve the "social learning goal". It is important to underline, that not all students relied on SC equally when making these decisions. The data analysis has shown that when we disregard the goal orientation, students' motivation does not seem to significantly depend on the presence of SC. However, when we analyze Mastery- and Performance-oriented students separately, differences between conditions (SC vs. No SC) become very noticeable. Mann-Whitney U Tests indicate that the motivation of Performance-oriented students was significantly impacted by SC, while for the Master-oriented students this effect was not present at all. Students in the *SC on Demand* group were not significantly different than any of the two groups but show behavior in the middle of the other two conditions. Table 1 summarizes the findings with median values of the corresponding measurements and the U-statistics.

**Table 1: Comparing the Performance and Mastery students in NoSC and SC Conditions**

| | Performance Oriented | | | Mastery Oriented | | |
|---|---|---|---|---|---|---|
| | **No SC** | **SC** | **U(p)** | **No SC** | **SC** | **U(p)** |
| n | 5 | 8 | | 15 | 11 | |
| Activities Accessed | 207 | 133.5 | 35 **(0.033)** | 156 | 204 | 57.5 (0.202) |
| Activities Revisited | 52 | 15.5 | 35 **(0.034)** | 25 | 60 | 45.5 (0.058) |
| # Sessions | 21 | 11.5 | 32 (0.092) | 14 | 22 | 45.5 (0.057) |
| # Days | 12 | 8 | 31 (0.122) | 10 | 13 | 49.5 (0.090) |

## 3 CONCLUSION

This study provides evidence that the effect of SC on students' motivation may vary depending on their Goal Orientations. It also demonstrates that a low social norm can make Performance-oriented students feel satisfied too early and disengage from learning. These findings lead to wider implications. When designing LADs, it is important to keep in mind that students' personal traits can render these interfaces individually ineffective and even harmful, especially when SC is utilized. We plan to continue this line of research by exploring various factors that may influence effectiveness of SC-enabled interfaces on individual students. A more general objective of this project is to develop a personalized framework for SC that is capable of supporting students' individual motivational needs.

## REFERENCES

Akhuseyinoglu, K., Milicevic, A., & Brusilovsky, P. (2022). Who are my peers? Learner- controlled social comparison in a programming course. *In: Proceedings of EC-TEL'2022*, (pp. 1–15).

Elliot, A. J. (2001). A 2× 2 achievement goal framework. *Journal of personality and social psychology, 80(3)*, 501-519.

Festinger, L. (1954). A theory of social comparison processes. *Human relations, 7(2)*, 117-140.

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice. *Data Driven Approaches in Digital Education: In: Proceedings of EC-TEL'2017*, (pp. 82-96).

Joshi, A., Molenkamp, B., & Sosnovsky, S. (2023). Student Perception of Social Comparison in Technology Enhanced Learning. *In Proceedings of EC-TEL'2023*, (pp. 118–132).

Manzoor, H., Akhuseyinoglu, K., Shaffer, C., & Brusilovsky, P. (2019). OpenDSA/Mastery Grids Exercise Interchange. *In Proceedings of the SPLICE Workshop colocated with SIGSCE'2019.*

Sosnovsky, S., Fang, Q., Vries, B., & Luehof, S. (2020). Towards adaptive social comparison for education. *In: Proceedings of EC-TEL'2020*, (pp. 421–426).

Teasley, S. D. (2017). Student facing dashboards: One size fits all? *Technology, Knowledge and Learning, 22(3)*, 377-384.

Wood, J. V. (1989). Theory and research concerning social comparisons of personal attributes. *Psychological bulletin, 106 (2)*, 231.

# A learning and teaching approach towards developing a maturity model for learning analytics in higher education

**C.A. Vonk[1]; R. Conijn[2]; A. Van den Beemt[1]**
[1]Eindhoven School of Education, Eindhoven University of Technology
[2]Human-Technology Interaction group, Eindhoven University of Technology
{c.a.vonk,m.a.conijn,a.a.j.v.d.beemt}@tue.nl

**ABSTRACT**: Higher education institutions increasingly apply Learning Analytics (LA). However, the upscaling of LA is limited. This poster aims to develop a LA maturity model to support a scalable implementation of LA. Existing models are often based on an institutional or technical approach. In contrast, the current model follows a learning and teaching approach and reflects both a descriptive and a prescriptive purpose. The model details the maturity on course, program, and institutional level which allows to focus on different aspects of the educational process. For the development of this maturity model a mini-Delphi method was used via interviews with multiple stakeholders. The model is currently further validated with an extended group of stakeholders via additional interviews.

**Keywords**: maturity model, higher education, learning analytics policy, learning analytics strategy, adoption at scale

## 1 INTRODUCTION

Learning Analytics (LA) is increasingly adopted in higher education beyond experiments of individual teachers [4]. The result is a range of LA interpretations and implementations, each with their own possibilities, challenges, and risks [3]. Within higher education institutions this requires among others, policy, quality assurance, allocation of funds and facilities, and professional development of educators and support staff. These requirements all aim at an effective use and benefit of LA for students, educators and management. However, despite the range of implementations, the practical use of scalable LA in higher education is still limited [3]. To monitor the institutional level of development of specific processes, such as the implementation of LA, and to show next steps in policy and practice, several researchers have proposed a maturity model as an instrument [8]. Within the field of LA already some maturity models have been proposed [2]. However, the need remains for a maturity model that supports scalable implementation of LA. Furthermore, existing models are all based on an institutional or technical approach, rather than a learning and teaching approach [5], while LA should start from a learning perspective [7]. The European maturity model for blended education (EMBED; [9]) can be considered as a model with learning and teaching focus. In this poster we aim to develop a maturity model following a learning and teaching approach, based on the EMBED model.

## 2 METHOD

The main guiding question for the current maturity model is: How does LA support teaching and learning activities? To develop the maturity model we followed the first four development stages proposed by [1], including scope, design, populate, and test as presented in Table 1. Furthermore, we adopted the basic design principles for maturity models, as well as the more specific design principles for descriptive and prescriptive models proposed by [6]. The model was built on literature and educational practice, by iteratively consulting relevant stakeholders at our university via a mini-Delphi approach [1]. For this, a total of 18 interviews were conducted with teachers, program directors, teacher support staff, policy makers and ICT support staff (see Table 1). More interviews will follow to further test construct validity.

**Table 1: Stages of the development of the maturity model.**

| Stage | Outcome | Validation (N = number of interviewees) |
|---|---|---|
| Inceptive | Develop maturity model as represented in the EMBED model [10] | Literature review, interviews (N=6) |
| Scope | Domain specific focus of LA policy and implementation to support teaching and learning | Interviews (N=1) |
| Design | Intended audience: internally oriented;<br>method of application: self assessment | Interviews (N=1) |
| Populate | The EMBED model translated to the LA domain, including identifying sub dimensions | Literature, Interviews (N=3) |
| Test round 1 | Check alignment with the local university practice | Interviews (N=7) |
| Test round 2 | Test for construct validity | Interviews (ongoing) |

## 3 THE MATURITY MODEL

The current maturity model is based on the EMBED model and consists of three levels: the course, the program, and the institutional level, with several dimensions and sub dimensions. The course level consists of four dimensions: course design process, course flexibility, course assessment, and course experience. The program level of three dimensions: program design, program flexibility, and program experience, and the institute level on 11 dimensions: institutional strategy, stakeholder involvement, stakeholder readiness, professional development, governance, finance, facilities, sharing, openness and dissemination, quality assurance, privacy and transparency, and fairness. Maturity is measured on three levels for all of these dimensions. The model provides a description for each of those maturity levels, which can be used to assess the level of maturity. An example of the model for one of the dimensions is presented in Table 2. The complete model is currently further assessed for construct validity via interviews with additional stakeholders, following current implementations of LA at our institution.

**Table 2: Maturity Model (one example dimension).**

| Dimension | Sub dimension | Level 1 | Level 2 | Level 3 |
|---|---|---|---|---|
| Course design process | Learning activities and sequences | **Exploratory**: Ad-hoc selection and integration of LA to support face-to-face and online | **Design-based**: LA are deliberately selected, integrated, and sequenced to support learning activities, based on | **Course cycle**: LA are deliberately selected, integrated, and sequenced to support learning activities, based on a design method or design principles. Quality |

| | | | | |
|---|---|---|---|---|
| | | learning activities. | a design method or design principles. | assurance processes are deliberately embedded to continuously improve course support. |

## 4 CONCLUSION

This study aimed to develop a maturity model for scalable implementation of LA in higher education. By building a learning and teaching based maturity model from existing models and frameworks for LA implementation in combination with an iterative design process with stakeholders, the model discusses a wide variety of aspects. We contend that this provides a more complete view of maturity within the institution. The model is in the process of validating within our institution. Eventually, this model could be further developed towards a comparative model, which could be used to benchmark LA implementation practices across higher education institutions. Potentially, the model could even be translated to other domains or strategical themes within higher education, making the model more generally applicable (cf.[7]).

## REFERENCES

[1] de Bruin, T., Rosemann, M., Freeze, R., & Kulkarni, U. (2005). Understanding the main phases of developing a maturity assessment model. In D. Bunker, B. Campbell, & J. Underwood (Eds.), *ACIS 2005 Proceedings - 16th Australasian Conference on Information Systems* (pp. 8–19).

[2] Freitas, E., Fonseca, F., Garcia, V., Ferreira, R., & Gasevic, D. (2020). Towards a maturity model for learning analytics adoption an overview of its levels and areas. *Proceedings of the IEEE 20th International Conference on Advanced Learning Technologies, ICALT 2020*, (pp 122-126). https://doi.org/10.1109/ICALT49669.2020.00059

[3] Hernández-de-Menéndez, M., Morales-Menendez, R., Escobar, C. A., & Ramírez Mendoza, R. A. (2022). Learning analytics: state of the art. *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 16(3), 1209–1230. https://doi.org/10.1007/s12008-022-00930-0

[4] Leitner P. Khalil M., Ebner. 2017. Learning analytics in higher education—a literature review. *Learning analytics: Fundaments, applications, and trends: A view of the current state of the art to enhance E-learning* 1–23.

[5] Marshall S. (2010). A Quality Framework for Continuous Improvement of E-learning: The E-learning Maturity Model. *Journal of Distance Education* 24, 1 (2010), 143–166

[6] Pöppelbuß, J., & Röglinger, M. (2011). What makes a useful maturity model? A framework of general design principles for maturity models and its demonstration in business process management. *ECIS 2011 Proceedings*, 28. https://aisel.aisnet.org/ecis2011/28

[7] Shum S., Ferguson R. (2012). Social learning analytics. *Educational Technology and Society 15, 3* (2012), 3–26. https://www.jstor.org/stable/jeductechsoci.15.3.3

[8] Tocto-Cano E., Paz Collado S., Linkolk López-Gonzales J., Turpo-Chaparro J. (2020). A systematic review of the application of maturity models in universities. *Information* 11, 10 466. https://doi.org/10.3390/info11100466

[9] Van Valkenburg, W., Dijkstra, W., de los Arcos, B., & Goeman, K. (2020). European Maturity Model for Blended Education. *Report published by European Association of Distance Teaching Universities*. empower.eadtu.eu

# Integration of SCAT and ENA in Quantitative Ethnography

**Daisuke Kaneko**
Hokusei Gakuen University, Japan
kaneko@hokusei.ac.jp

**Ayano Ohsaki**
Shinshu University, Japan
aohsaki@ohsaki-lab.net

**ABSTRACT**: In quantitative ethnography, epistemic network analysis (ENA) stands out as a method capable of uncovering human cognition embedded in qualitative data. ENA requires the coding of qualitative data, presenting challenges related to the need for consistency and validity between the data and analytical results, as well as a demand for transparency in the analytical process. Therefore, we propose the use of steps for coding and theorization (SCAT) for explicit code development. We analyzed the interview data using SCAT, generated codes from this process, applied the codes to the ENA analysis, and discussed our results and future research challenges.

**Keywords**: Quantitative Ethnography, ENA, SCAT, Coding, Qualitative Data Analysis

## 1 BACKGROUND

Quantitative ethnography (QE) is a methodology used to analyze large-scale qualitative data. As a representative QE technique, epistemic network analysis (ENA) identifies and quantifies relationships between elements in coded qualitative data to reveal how human cognition is represented (Shaffer 2017). The ENA coding process comprises two stages: code development, where codes are generated based on insights gleaned from qualitative data, and code assignment, where the segmented data are verified to contain relevant codes. This study specifically focuses on code development.

Shaffer and Ruis (2021) provided guidelines for good coding practices and addressed coding issues, emphasizing the crucial linkage between data and analysis in terms of consistency and fairness. Otani (2019) highlighted that analyst reflectivity and reader falsifiability are essential for qualitative research data analysis, underlining the importance of transparency and comprehensibility in analytical procedures. To address these issues, explicit coding must be performed deliberately. We utilized the qualitative data analysis method, steps for coding and theorization (SCAT), to ensure transparent code generation (Kaneko and Ohsaki 2023, Ohsaki and Kaneko in press). SCAT, known for its explicit and formalized procedure, is suitable for beginners and has been used to analyze relatively small datasets (Otani 2015, 2019). We applied SCAT to analyze interview data from an experienced wind orchestra instructor and subsequently developed codes for ENA analysis.

## 2 METHOD

SCAT involves a four-step coding process, wherein segmented data are described within a matrix, and codes are assigned in the following order: Step <1>: Extract noteworthy words or phrases from the text. Step <2>: Paraphrase <1>. Step <3>: Put extra-textual concepts that account for <2>. Step <4>: Emerge themes or concepts considering the context. After that, create a <story-line> using <4> , and finally, derive a <theory> from the story-line. SCAT procedure ends here, but in this study, synthesize all theories into a single framework and assign multiple codes to each theory being conscious of the overall context. The codes and their definitions are compiled in a codebook. Following this procedure makes code development explicit, thereby enhancing the transparency of the coding process.

Meanwhile, we developed a system designed to support individual practice for wind-instrument players, especially beginners, by visualizing pitch and volume. To evaluate the effectiveness of this system, we conducted trials with wind orchestra students. As a preliminary step, we conducted a semi-structured interview with an instructor from the wind orchestra. This study aimed to clarify the instructor's perspective on musical instruction and the implemented system. Throughout this investigation, we attempted to analyze the interview data using a method previously described.

## 3 RESULTS AND FUTURE RESEARCH

The sample results from the analysis of the interview data using SCAT are listed in Table 1. A total of 361 utterances were analyzed, leading to the derivation of 12 theories. Based on these theories, eight codes were assigned (see Table 2). Subsequently, we assigned these codes to the interview data and conducted an ENA analysis, resulting in the network diagram shown in Figure 1. Notably, a robust connection was observed between instruction and issues, awareness, and skill development. From these results, it can be suggested that the instructor's perspective on instruction is not directly related to technology. To verify whether this relationship changes with the future introduction of the system, post-test interviews will be necessary.

Our results illustrate the potential of SCAT in ensuring transparency in the coding process. However, several issues must be addressed. While our focus was on code development, we also intend to enhance the transparency of code assignment and address the challenge of simultaneously employing methodologies with differing epistemologies. Qualitative methods have interpretive epistemologies, deriving meaning through text interpretation, whereas quantitative methods embrace positivist epistemologies, posing the risk of inherent discrepancies in analytical findings. In this context, whether the four-step coding in SCAT is merely used as a process of clarification or a means to transcend epistemological differences requires further discussion.

## ACKNOWLEDGMENT

This research is supported by JSPS KAKENHI (Grant Numbers 21K12178 & 23K11357), 2023 Hokusei Gakuen Univ. Specific Research Grant and Editage (www.editage.jp) for English language editing.

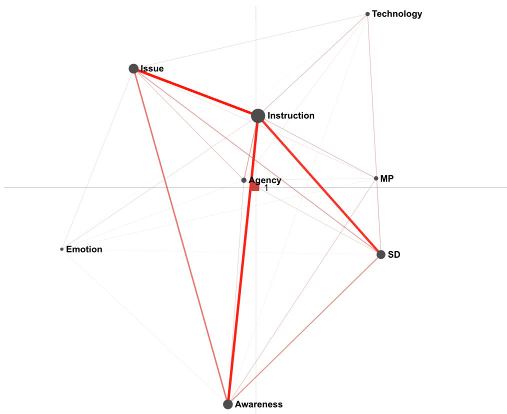



**Figure 1. Network diagram resulting from ENA analysis**

**Table 1. Sample results from analysis of interview data using SCAT**

| No | Speaker | Text (English) | Step <1> | Step <2> | Step <3> | Step <4> |
|---|---|---|---|---|---|---|
| 61 | Interviewer (Int.) | How you give instructions and such. | instructions | | | |
| 62 | Instructor (Inst.) | Well, basically, it's not so much about telling them what to do, but rather having the students try things out for themselves. | having the students try things out | Student-Led Approach | Student-Entrusted Practice Methods/ Self-Authority in Practice | Perception of Students as Autonomous Learners |
| 63 | Int. | I see. | | | | |
| 64 | Inst. | And, of course, there are beginners who just can't do it, so for those students, I'll show them, like, 'Try playing it this way,' and guide them one-on-one instruction. But after that, the seniors take over and I mostly leave it to them. | beginners/one-on-one instruction/ seniors/ leave it | Beginner/ Supportive Coaching Method/ Initially Only/ Senior Students' Autonomy/ Apprenticeship-style/ Delegating | Detailed Guidance from Instructors for Instrument Beginners/ Autonomous Instruction by Seniors/ Shift in Instructional Responsibility | Transition of Instructional Responsibility to Experienced Senior Instrumentalists |
| 65 | Int. | But even when you say one-on-one instruction, there's no separate place to produce sound, right? (snip) | no separate place to produce sound | Lack of Individual Practice Environment | Lack of Individual Instruction Environment | |
| <Story-line> | | In practice sessions led by instructors who view students as 'Perception of Students as Autonomous Learners,' there is an observed 'Transition of Instructional Responsibility to Experienced Senior Instrumentalists.' However, … (snip) | | | | |
| <Theory> | | (1) In practice sessions led by instructors who view students as 'Perception of Students as Autonomous Learners,' there is an observed 'Transition of Instructional Responsibility to Experienced Senior Instrumentalists.'<br>(2) There remains a 'Necessity of Basic Individual Instruction for Beginners,' necessitating 'Instructor-Led Guidance for Solving Performance Challenges of Beginners.' | | | | |

**Table 2. Coding table**

| Code | Definition |
|---|---|
| Agency | The individual's ability to control their own actions and act (perform or practice) voluntarily. |
| Instruction | Matters related to the instructional strategies of a leader, including specific ways of instruction and thoughts about teaching. |
| Musical Perspective | An individual's views and approaches to music, encompassing values, preferences, interpretation methods, and understanding of music's composition. |
| Skill Development | Refers to the growth and improvement of musical skills and abilities. Also includes references to beginners and experienced individuals. |
| Emotion | Pertains to individuals' emotions, such as motivation in students and the enjoyment of performing music. |
| Awareness | Refers to mentions of the current situation, excluding mentions to problems or challenges. |
| Issue | Mentions related to problems, challenges, and conflicts or disputes. |
| Technology | Mentions related to the system and perceptions of the system |

## REFERENCES


Kaneko, D., Ohsaki, A. (2023). An Approach to Quantitative Ethnography with SCAT: the Use of Qualitative Data Analysis Method, In: Irgens, G.A., Knight S. (eds) *the Fifth International Conference on Quantitative Ethnography: Conference Proceedings Supplement*, pp.61-62.

Ohsaki, A., Kaneko, D. (in press). Computer-Supported Code Generation Using Stepwise Coding. In *Proceedings of 14th International Conference on Learning Analytics and Knowledge*, LAK24.

Otani, T. (2015). SCAT: Steps for Coding and Theorization, https://www.educa.nagoya-u.ac.jp/~otani/scat/index-e.html, last accessed 2024/1/29.

Otani, T. (2019). *Paradigm and Design of Qualitative Study: From Research Methodology to SCAT*. The University of Nagoya Press, Nagoya, Aichi, Japan. (In Japanese).

Shaffer, D.W. (2017). *Quantitative Ethnography*. Cathcart Press, Madison, WI.

Shaffer, D.W., Ruis, A.R. (2021). How We Code. In: Ruis, A.R., Lee, S.B. (eds) *Advances in Quantitative Ethnography*. ICQE 2020, CCIS 1312, pp.62-77. Springer, Cham.

# Effects of Instructional Design and Design Preferences on Cognitive Load, Problem Solving, and Learning Outcomes in a Computer-based Office Simulation

**Sabrina Ludwig**
Mannheim Business School, University of Mannheim, Mannheim, Germany
sabrina.ludwig@uni-mannheim.de
**Andreas Rausch**
Mannheim Business School, University of Mannheim, Mannheim, Germany
rausch@uni-mannheim.de
**Michelle Taub**
Department of Learning Sciences and Educational Research, University of Central Florida, United States of America
Michelle.taub@ucf.edu

**ABSTRACT**: This study examines the effects of the two instructional designs — 'direct instruction before problem solving (DI-PS)' and 'problem solving before direct instruction (PS-DI)' — on cognitive load, problem-solving processes, performance, and learning outcomes in an office simulation. Business undergraduate students (n = 81) were assigned to either the DI-PS or PS-DI condition. Educational data mining techniques such as regressions were employed to examine self-regulated learning (SRL) in an office simulation. Different learning strategies (such as reading in the reference books) across the groups could be identified using log data and applying learning analytics. Results showed that the DI-PS group performed better in problem solving, possibly due to significantly lower in-scenario and retrospective cognitive load. Only minor differences can be reported in procedural knowledge scores. The DI-PS group also reported higher satisfaction with its instructional sequencing. These findings are consistent with existing literature on instructional design and instructional preferences. Implications for instructional designers regarding task complexity were derived.

**Keywords**: Instructional Design, Problem Solving, Direct Instruction, Cognitive Load, Experimental Study, Log File Analysis, Simulation-Based Learning

## 1. INTRODUCTION

Computer-based simulations provide several benefits to learners, including the ability to engage in self-regulated learning (Azevedo & Gašević, 2019; Munshi et al., 2018) and inquiry-based problem solving to develop domain-specific problem-solving performances (Chernikova et al., 2023). While problem-solving (PS) is characterized as active and challenging, it can also result in a cognitive load that is too high (Likourezos & Kalyu, 2017). Prior worked examples (e.g., video tutorials) as direct instruction (DI) can reduce cognitive load by providing the necessary background knowledge and allowing learners to acquire a task-related schema (Sweller, 1994; van Merrienboer & Sweller, 2005). According to Loibl et al. (2020), several studies show that learners who receive direct instruction prior to problem-solving (DI-PS) tend to acquire more procedural knowledge compared to those who do not receive prior instruction (e.g., Chen et al., 2015; van Gog et al., 2011). Conversely, PS-DI enhances learning transfer by activating prior knowledge and highlighting gaps during problem solving ('productive failure'; Kapur, 2014). DeCaro and Rittle-Johnson (2012) demonstrate that PS-DI supports conceptual understanding better than DI-PS. Additionally, individual instructional preferences play a role in learning, with some learners benefiting more from PS-DI while others favor DI-PS. This leads to inconsistent empirical findings on these instructional designs (Chen & Kalyuga, 2020). Moreover, the impact of instructional designs (DI-PS and PS-DI) in the business field has not yet been explored, and whether differences in problem-solving processes and cognitive load exist should be examined. Hence, this study aims to assess the effects of both instructional designs on problem-solving performance and knowledge outcomes in a business simulation. Additionally, it sheds light on the problem-solving processes and cognitive load of learners with business background. Against this background, the following research questions (RQ) are addressed:

1) How do instructional design (DI-PS vs. PS-DI) and instructional design preferences explain problem-solving performance and learning outcomes?
2) How does instructional design (DI-PS vs. PS-DI) affect problem-solving behaviour and cognitive load?
3) How does problem-solving behaviour predict problem-solving performance and learning outcomes within the DI-PS and PS-DI groups?

## 2. METHOD

An experimental study employing a pre-post-test design with 81 students was conducted. The students engaged in the authentic business case 'supplier selection' within an office simulation equipped with typical office tools like an email client, PDF viewer, spreadsheet program, reference book, calculator, and notepad (Figure 1). The students were randomly assigned to either DI-PS or PS-DI. The study was structured into four phases (1-4): Following an introduction by the instructors (phase 1), participants in both conditions completed questionnaires (covering demographics and instructional preferences) and a knowledge pretest (phase 2). In phase 3, all participants engaged in an onboarding scenario to familiarize themselves with the simulation. In the DI-PS condition, participants viewed a 6-minute direct instruction video (DI, see Figure 2) before completing a 40-minute problem-solving task (PS, Figure 1). In the PS-DI condition, participants started with the PS task, followed by the DI. In phase 4, all participants completed a knowledge posttest to assess their learning outcomes and completed a retrospective instructional preference questionnaire and a feedback questionnaire, including items to measure cognitive load.

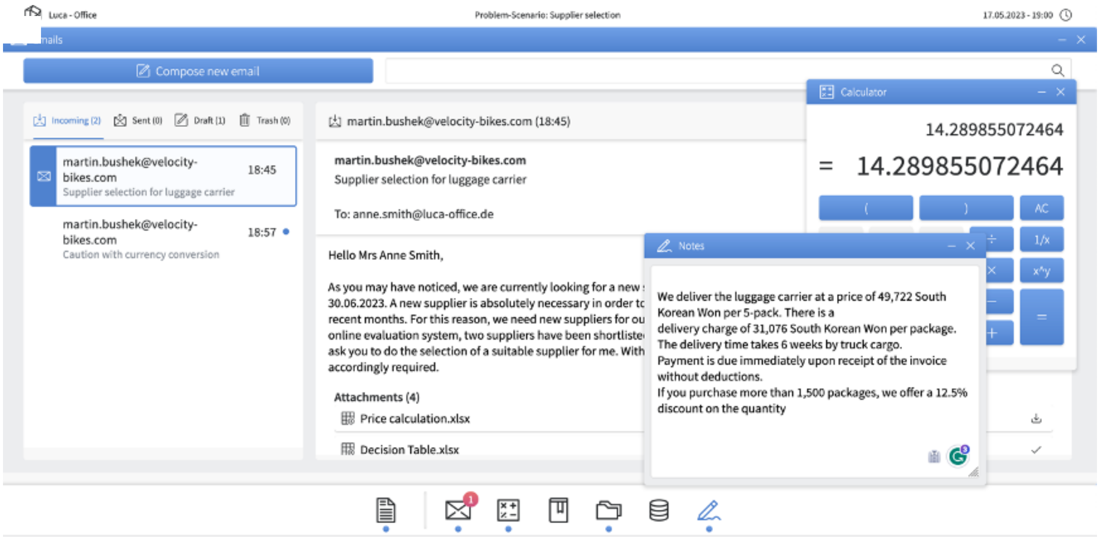



**Figure 1: Office simulation with an email client, the task assignment, a notepad, and a calculator**

In assessing problem-solving performance, students' responses and calculations in the problem-based scenario (PS) were evaluated based on a domain-specific performance model with eleven criteria (such as correct argumentation). To examine conceptual and procedural learning outcomes, differences (Δ) between pre- and posttest scores were used, and mixed ANOVAs as well as ANCOVAs were performed. To assess the instructional preferences, the participants completed questionnaires before and after the scenarios. Cognitive load was measured during the PS (in-scenario experience sampling) and afterwards by using a questionnaire. To explore problem-solving processes, log data (including mouse clicks and keystrokes) were collected during the problem-solving (PS) scenario. Similar subsequent actions were then aggregated, and behavioral indicators (e.g., note-taking or reading reference books) were derived. In addition, we used Educational Data Mining techniques such as regressions to examine the relationship between behavioral indicators and problem-solving performance as well as learning outcomes.

## 3. Findings

Regarding RQ1, the DI-PS group achieved higher mean scores ($M = 22.13$ of 31 in total) for problem-solving performance in the problem-solving task (PS) than the PS-DI group ($M = 19.24$). Especially, DI-PS group significantly outperformed the PS-DI in two scoring items 'correct argumentation for supplier A' and 'correct argumentation against supplier B' ($p = .01$ and $p = .007$, respectively). Moreover, the mixed ANOVA shows that the PS-DI group gained significantly higher conceptual knowledge than the DI-PS group (MPS-DI = 1.54 credits; MDI-PS = .93 credits; $p = .05$). Additionally, both groups show significantly higher scores in the post-tests compared to the pre-tests for both conceptual and procedural knowledge. Significant interaction effects (groups x pre-/posttest) were only found in procedural knowledge. Two ANCOVAs with instructional preferences as covariate shows that instructional preference had a significant effect on conceptual but not on procedural knowledge gain in both groups. 80% of the DI-PS group and 83% of the PS-DI group stated that they preferred starting with learning theories (DI) before tackling complex and practical PS tasks. Regarding RQ2, examining the log-file data and deriving behavior indicators reveal that the PS-DI group significantly exhibited a higher frequency of activity in reference books (M = 4.5) than the DI-PS group (M = 1.3). Cohen's d of .83 shows a high effect size. Regarding RQ 3, regression analyses demonstrate that engaging with both spreadsheets 'Decision table' and 'Price Calculation' in the PS task had a substantial positive impact on problem-solving performance for both groups. Group DI-PS also showed that 'Price Calculation' ($R2$ = .10, $F(1, 38) = 4.15$, $p = .05$) is a good predictor for conceptual knowledge gains, as well as 'Reading Documents' ($R2$ = .11, $F(1, 38) = 4.77$ $p = .04$). Furthermore, the PS-DI group shows significantly higher cognitive load (MDI-PS = 2.10 vs. MPS-DI = 2.50, $p = .03$, moderate effect size: d = .53) by rating the PS task during and after it as difficult.

## 5. Discussion and conclusion

The study adds to the knowledge of the influence of instructional sequencing and instructional preferences on problem-solving performance, learning outcomes, cognitive load, and problem-solving processes. Unsurprisingly, the DI-PS group showed higher problem-solving performance than the PS-DI group, which is consistent with the findings of Dubovi (2018). The manipulation of prior knowledge by providing a video as direct instruction reduced in-scenario intrinsic cognitive load (Sweller, 1994) and, eventually, resulted in higher problem-solving performance. While previous studies suggest that PS-DI may facilitate conceptual learning (Kapur, 2014), significant group differences in conceptual learning outcome were also found in this study. Regarding problem-solving processes, the DI-PS group showed that spreadsheet calculations and reading documents (including reference books) positively affect problem-solving performance, which is in line with Yamada et al. (2017). Furthermore, the higher satisfaction with the instructional sequence among participants in the DI-PS condition resonates with existing literature. This preference could be attributed to the teaching methods and learning culture, as explained by Meyer (2014). Implications for instructional designers will be outlined in our conference contribution.

## References


Azevedo, R., & Gašević, D. (2019). Analyzing multimodal multichannel data about self-regulated learning with advanced learning technologies: Issues and challenges. Computers in Human Behavior, 96, 207-210.

Chen, O., Kalyuga, S., & Sweller, J. (2015). The worked example effect, the generation effect, and element interactivity. Journal of Educational Psychology, 107(3), 689.

Chen, O., & Kalyuga, S. (2020). Exploring factors influencing the effectiveness of explicit instruction first and problem-solving first approaches. European Journal of Psychology of Education, 35(3), 607–624. https://doi.org/10.1007/s10212-019-00445-5

Chernikova, O., Holzberger, D., Heitzmann, N., Stadler, M., Seidel, T. & Fischer, F. (2023). Where salience goes beyond authenticity. A meta-analysis on simulation-based learning in higher education. Zeitschrift für Pädagogische Psychologie, online first, https://doi.org/10.1024/1010-0652/a000357

DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. Journal of experimental child psychology, 113(4), 552-568.

Dubovi, I. (2018). Designing for online computer-based clinical simulations: Evaluation of instructional approaches. Nurse Education Today, 69, 67-73.

Kapur, M. (2014). Productive failure in learning math. Cognitive science, 38(5), 1008-1022.

Loibl, K., Tillema, M., Rummel, N., & van Gog, T. (2020). The effect of contrasting cases during problem solving prior to and after instruction. Instructional Science, 48, 115-136.

Likourezos, V., & Kalyuga, S. (2017). Instruction-first and problem-solving-first approaches: Alternative pathways to learning complex tasks. Instructional Science, 45, 195-219.

Meyer, E. (2014). The culture map: Breaking through the invisible boundaries of global business. Public Affairs.

Munshi, A., Rajendran, R., Ocumpaugh, J., Biswas, G., Baker, R. S., & Paquette, L. (2018, July). Modeling learners' cognitive and affective states to scaffold SRL in open-ended learning environments. In Proceedings of the 26th conference on user modeling, adaptation and personalization (pp. 131-138)

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. Learning and Instruction, 4(4), 295–312. doi:10.1016/0959-4752(94)90003-5.

Van Merrienboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. Educational Psychology Review, 17(2), 147–177. doi:10.1007/s10648-005-3951-0.

Van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. Contemporary Educational Psychology, 36(3), 212-218.

Yamada, M., Shimada, A., Okubo, F., Oi, M., Kojima, K., & Ogata, H. (2017). Learning analytics of the relationships among self-regulated learning, learning behaviors, and learning performance. Research and Practice in Technology Enhanced Learning, 12(1), 1-17.

# Towards the design of an interactive study planning tool integrating AI-based feedback and learning analytics

**Johannes Wagner-Schirrmeister**
Ruhr University Bochum
johannes.wagner-schirrmeister@rub.de

**Rene Roepke**
RWTH Aachen University
roepke@cs.rwth-aachen.de

**Rike Carpantier, Maren Scheffel**
Ruhr University Bochum
rike.carpantier@rub.de, maren.scheffel@rub.de

**ABSTRACT**: The challenges of study planning lie in available strategies and support resources when students divert from recommended plans while still aiming to successfully graduate. Resources like exam regulations and module handbooks are of limited use due to their static nature, failing to provide insights into successful study paths and choices of past students. To this end, artificial intelligence and learning analytics can be utilized to support student planning with feedback, indicators, and recommendations. This poster presents the design and implementation of an interactive study planning tool employing artificial intelligence-based feedback and data-driven insights derived from using process mining on curriculum data.

**Keywords**: Study Planning, Feedback, Artificial Intelligence, Process Mining

## 1 INTRODUCTION

In response to the challenge of long-term study planning, students in higher education have access to various resources for support and guidance. Common services like mentoring programs or advisory offices may provide support either upon starting university or when times get tough. Usually, essential online resources include exam regulations and module handbooks of respective study programs. While at first, a recommended study plan proposed in the exam regulations may provide a useful starting point for planning the first semester, its value decreases significantly when students start to divert from recommended plans (e.g., by failing courses they must retake in the future or when they do not want to adhere to the recommended workload due to personal circumstances or other external factors). To address this issue, different research projects explore interactive study planning tools for students (e.g., Judel et al., 2023; Weber et al., 2022; Hirmer et al., 2022). Methods of artificial intelligence (AI) can be used to provide planning feedback as well as conformance checking to allow for valid alternative study plans. By analyzing students' paths through study programs, insights on successful paths can be gained along with indicators of choices increasing the probability of failure. To this end, predictive analytics can be used to provide success probabilities and to forecast study success. Related work on process mining with curriculum data demonstrates how to detect study paths in order to recommend follow-up courses (Schulte et al, 2017). Further work has been summarized by Bogarin et al. (2018). Not yet found in literature are approaches combining process mining and rule-based AI for study planning. To this end, we aim to investigate how process mining, rule-based AI, and learning analytics (LA) can be combined to support comprehensive study planning for students. In this poster[1], we introduce the design and implementation of our prototype.

## 2 PRE-STUDY TO COLLECT REQUIREMENTS

To gain an initial understanding and collect requirements, a pre-study of prospective users, their behavior, interests, needs, and requirements was conducted. Its results formed the basis for the initial design and prototyping process of an interactive study planning tool. The survey was answered by n=674 students from three universities, 50.7% of which fully completed the survey. Most notably, the results show that students overwhelmingly organized their planning according to recommended study plans. While a majority cited exchange with fellow students from their cohort as a regular factor, only a minority frequently based their decisions on prior student experience (offered by study program websites, student bodies, or alumni). Actual guidance and advisory services were least frequently utilized. A majority already used integrated digital systems (e.g., CMS tools) as information sources and planning aids. However, these tools so far neither enable students to make individual planning decisions nor do they allow students to reflect on plans or the validity of plans and possible choices. This clearly demonstrates a gap: So far, there are no study planning aids, no LA and AI-based feedback tools which enable students to make informed individual study planning decisions that are based on their specific needs and requirements as well as on knowledge of prior student experience.

## 3 DESIGN AND IMPLEMENTATION OF THE TOOL

*User Interface:* The tool is implemented as a single-page application using an Angular-based frontend and a node.js backend. Designed as separate micro-services, the application connects to a symbolic, rule-based AI component for feedback on planning actions and a LA backend for data-driven insights. Recommended plans taken from exam regulations are used as an initial starting point, based on which students may select a start semester and then adjust all modules according to their current study progress as well as plans for upcoming semesters. Additional columns for future semesters can be added as needed, allowing students to plan beyond the expected standard period of study.

*Feedback through AI:* The study program model with its rules and regulations forms the basis of the AI-based feedback component. Depending on the availability of a machine-readable model and its rules, manual conversion of exam regulations into formal notation is required. Concerning a symbolic, rule-based AI approach, the rules and regulations of a study program are translated into a custom notation, suitable to be evaluated using formal logic calculus. Extending the notation of event calculus with eventualities and concepts of actions performed during planning (a sub-discipline of AI), we allow for evaluating past events as well as newly planned events in order to provide feedback on students' decisions and possible rule violations past, present and future. Thus, a student-generated study plan with successfully completed past modules and partially arranged modules for current and future semesters can be evaluated while explainable feedback allows students to adapt their plans before triggering the next evaluation.

[1] The poster will be publicly available after the conference, see DOI: 10.18154/RWTH-2024-00513

*Support through recommendations and analytics:* While study program regulations can be checked immediately through rule-based AI, we also aim to support the planning process itself through data-driven insights. Here, LA can be implemented to provide insights into basic indicators (e.g., module attendance and pass/fail rates) as well as advanced indicators (e.g., success probability based on students' prior study profile). Even more advanced LA functionality can be provided, e.g., through student cohort analyses, providing students with insights into the current progress of immediate peers in the same semester. To achieve this, process mining using CMS data is utilized to gain a deeper understanding of study paths by discovering process models and comparing them to intended processes, i.e. recommended study plans and partial module orders based on recommendations in module handbooks. Using diagnostic measures, the conformance and performance of discovered models can be computed. Using predictive approaches, we can utilize process models to formulate recommendations, e.g., selection of modules for upcoming semesters or suitable planning actions when moving a particular module.

Concerning the use of AI to generate feedback and process mining to derive recommendations from past study paths, we must consider that respective designs within a study planning tool may have an immediate impact on students. Different implementation variants should therefore be considered. Furthermore, provisioning data-driven insights requires a basic level of trust in the system. Accordingly, students should be able to control how and when recommendations are provided. To support individual, autonomous planning, feedback should be configurable, e.g. deactivated when it is not wanted or needed. We aim to provide students with sensible options for adjusting the tool to their preferences and to evaluate implementations of AI and process mining accordingly. To this end, we employ a formative, criterion-based evaluation framework, encompassing perspectives of *usability*, *acceptance*, *ethics*, *privacy*, *pedagogy*, and *improvement potential*, as an integral part of the design and implementation process. In future work, we plan to properly present our evaluation framework as well as results of ongoing evaluation studies conducted with students and relevant stakeholders, combining iterations of user tests with the collection and analysis of survey and interview data.

## REFERENCES

Bogarín, A., Cerezo, R., & Romero, C. (2018). A Survey on Educational Process Mining. *WIREs Data Mining and Knowledge Discovery 8*, 1 (2018), e1230.

Hirmer, T., Etschmann, J., & Henrich, A. (2022). Requirements and Prototypical Implementation of a Study Planning Assistant in CS Programs. *2022 International Symposium on Educational Technology (ISET)*. IEEE, New York, 281–285.

Judel, S., Roepke, R., Azendorf, M. & Schroeder, U. (2023). Supporting Individualized Study Paths Using an Interactive Study Planning Tool. *21. Fachtagung Bildungstechnologien (DELFI)*. Gesellschaft für Informatik e.V., Bonn, 225–230.

Schulte, J., Fernandez de Mendonca, P., Martinez-Maldonado, R., & Buckingham Shum, S. (2017). Large Scale Predictive Process Mining and Analytics of University Degree Course Data. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK '17)*. ACM, New York, NY, USA, 538–539

Weber, F., Schrumpf, J., Dettmer, N., & Thelen, T. (2022). Web-Based Recommendation System for Higher Education: SIDDATA: History, Architecture and Future of a Digital Data-Driven Study Assistant. *International Journal of Emerging Technologies in Learning (iJET) 17*, 22, 246–254.

# How higher education students use student-facing learning analytics for self-regulated learning: A systematic review.

**Theo C.C. Nelissen[1], Esther J. van der Stappen[1], Milou van Harsel[1] and Anique B.H de Bruin[2]**
[1]Avans University of Applied Sciences, [2]Maastricht University,
tcc.nelissen@avans.nl

**ABSTRACT** POSTER: Student-facing learning analytics (SFLA) offer new possibilities to formulate and deliver external feedback in support of self-regulated learning (SRL). But students seem to lack the necessary competencies to make sense of SFLA for SRL and how students actually engage with SFLA for SRL in authentic higher education settings is under-researched. This systematic literature review, following the PRISMA framework, explores how students use SFLA for SRL and what student characteristics are beneficial. A search with a broad set of keywords retrieved 3.487 records, published between 2013 and 2023. The selection of 36 articles described an SFLA intervention, reported process data and a measure of effect on SRL. In this poster, we present our preliminary findings.

**Keywords**: Self-regulated learning, systematic review, student-facing learning analytics.

## 1 INTRODUCTION

In recent years, higher education has increasingly offered students autonomy in shaping their study paths. When receiving more autonomy, self-regulated learning (SRL) is a critical factor to learn effectively (Boelens et al., 2017). SRL features a continuous flow of information between a student engaging in learning and metacognitive monitoring, in turn triggering action or regulation (de Bruin et al., 2020). However, students often lack necessary competencies to accurately regulate their learning (Viberg et al., 2020). Students who lack self-regulatory skills are less capable to generate internal feedback and rely more on external sources of feedback. Educational data-related technologies, like learning analytics (LA), offer new possibilities to improve SRL by offering personalized feedback, delivered in time, at scale and (partly) automated (Pardo et al., 2017). LA can be considered as a source of information (external feedback) for students to transform it into internal feedback to support SRL (Nicol, 2021; Viberg et al., 2020). If the data or feedback is directly presented to students, without conveyance by, for instance, an instructor, a LA system (e.g. LAD) is considered student-facing learning analytics (SFLA) (Bodily & Verbert, 2017). But not all students are adequately prepared to make sense of the information presented (Jivet et al., 2020). How students actually engage with SFLA in real-world settings and how this influences their SRL is still under-researched (Jivet et al., 2020; Pardo et al., 2017). This systematic review adds to the knowledge base on how students use SFLA for SRL in authentic higher education settings.

## 2 RELATED WORK

Several models of SRL exist, describing SRL as a cyclical process consisting of different related phases or subprocesses (Panadero, 2017). This research uses the COPES-model of SRL by Winne and Hadwin as guidance. This model is often used in research relating to computer supported learning environments and makes external evaluations explicit (Panadero, 2017). On the subject of SRL and LA, several reviews have been conducted. The literature review on student-facing learning analytics dashboards and recommender systems by Bodily and Verbert (2017) expressed the need for experiments to determine the effect of these systems on student behavior, achievement, and skills, including methodologies to examine student use. They also conclude that student use of study-data presented in a system is generally low. Bodily and Verbert did not specifically focus on SRL, but in most cases, stimulating awareness or reflection was the main intended purpose of student-facing systems. The state of applications of LA to measure and support SLR in online learning environments is highlighted by Viberg et al. (2020). This review found little evidence for improvements in learning outcomes and in learning support. LA was also not used widely. The insight in whether, to what extent and how data is used by students was limited. Viberg et al. did not include subprocesses of SRL in the search string, possibly excluding relevant results. Heikkinen et al. (2023) looked at applied channels and methods and how studies evaluated effects. Trace data was the most used method. A positive, measurable impact on learning was reported by 46 percent of the studies. However, students' self-evaluation did not always align with the impact seen in trace data.

Analyzing how students use LADs, is an important element of determining the effectiveness of LADs, but most articles do not report on student use of LADs (Bodily et al., 2018). Lim et al. (2021) call for more in-depth information from a student's perspective, to make firmer conclusions about the link between the intervention and the outcomes. To gain insight into how HE students use SFLA for SRL in authentic educational settings, this systematic review will address which phase(s) of SRL is/are supported by SFLA (Q1), how students interact with SFLA for SRL in authentic educational settings (Q2) and what student characteristics are considered beneficial (Q3).

## 3 METHOD

This systematic literature review follow the PRISMA guidelines. Because several authors mention that how students actually engage with LA in real-world settings is under-researched, we decided to include a more broad array of search terms at the start, acknowledging a) the variety in SRL models and related terminology and b) the different domains and technology that student facing, LA interventions can relate to. We included keywords relating to subprocesses of SRL, the learning analytics process model, LA, Open Learner Models and Educational data mining. The full list of keywords will be included in the poster. We included peer-reviewed English literature published between 2013 and March 2023 (moment of search). In line with Jivet et al. (2018) the concept of student-facing is incorporated as an inclusion criterium instead of in the search keywords. Eligibility criteria were: 1) SFLA intervention for HE students, without synchronous guidance by others, 2) process data on HE students' use of SFLA for SRL in authentic educational settings and 3) SRL or a subprocess of SRL is studied as outcome variable. The literature search included Education Research Complete ((educational)science), PsycINFO (SRL), IEEE Xplore (computer science) and ACM Digital Library (LAK full papers). A total of 3.487 records were identified. After deleting duplicates, 3.220 abstracts were screened resulting in 173 reports for full text screening. The eligibility criteria were tested on a small selection of articles and discussed by the research team. During both the abstract screening (n=3.220) and the full text screening (n=173), the first and a second author separately screened a random subset of 10 percent. Conflicts were discussed and resolved before screening the remaining records. A set of 36 articles matched the eligibility criteria.

## 4 PRELIMINARY FINDINGS

Currently, the final set of papers is being analyzed and interpreted. Since some of our preliminary conclusions are likely to be of interest to the LAK community, we present them in this poster. Monitoring learning, creating awareness, and fostering reflection are processes regularly mentioned as outcome of the use of SFLA. This is in line with previous research by Bodily & Verbert (2017). Among the included studies, there are several studies describing processes related to SRL without explicitly mentioning SRL in the abstract or title. This seems to support our choice to formulate a broad set of keywords in our search. Digital trace data are a promising source of information on how students use SFLA. The relevance of trace data is dependent on the way educational technology, educational theory and educational practice are intertwined. Relevant trace data can both give insight into how students use SFLA for SRL and serve as an outcome measure of SRL. In-depth information on how individual students use SFLA for SRL is often absent. A limitation we found in the selected studies, is that whenever process data is collected, it is often collected from students who indeed made use of the system, not giving insight into the barriers students not using the system (anymore) encountered.

## REFERENCES

Bodily, R., Ikahihifo, T. K., Mackley, B., & Graham, C. R. (2018). The design, development, and implementation of student-facing learning analytics dashboards. Journal of Computing in Higher Education, 30(3), 572–598. psyh. https://doi.org/10.1007/s12528-018-9186-0

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)

Boelens, R., De Wever, B., & Voet, M. (2017). Four key challenges to the design of blended learning: A systematic literature review. Educational Research Review, 22, 1–18.

de Bruin, A. B. H., Roelle, J., Carpenter, S. K., Baars, M., & EFG-MRE. (2020). Synthesizing Cognitive Load and Self-regulation Theory: A Theoretical Framework and Research Agenda. Educational Psychology Review, 32(4), 903–915. https://doi.org/10.1007/s10648-020-09576-4

Heikkinen, S., Saqr, M., Malmberg, J., & Tedre, M. (2023). Supporting self-regulated learning with learning analytics interventions – a systematic literature review. Education & Information Technologies, 28(3), 3059–3088. https://doi.org/10.1007/s10639-022-11281-4

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.

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.

Nicol, D. (2021). The power of internal feedback: Exploiting natural comparison processes. Assessment & Evaluation in Higher Education, 46(5), 756–778.

Panadero, E. (2017). A Review of Self-regulated Learning: Six Models and Four Directions for Research. Frontiers in Psychology, 8, 422. https://doi.org/10.3389/fpsyg.2017.00422

Pardo, A., Poquet, O., Martinez-Maldonado, R., & Dawson, S. (2017). Provision of Data-Driven Student Feedback in LA & EDM. In Handbook of Learning Analytics (pp. 163–174). SOLAR.

Viberg, O., Khalil, M., & Baars, M. (2020). Self-Regulated Learning and Learning Analytics in Online Learning Environments: A Review of Empirical Research.

# (De)motivating Low-Performing Students with AI-Based Negative Feedback: Do Incorrectness Cues Matter?

**Authors: Marlene Steinbach[1], Johanna Fleckenstein[2], Livia Kuklick[1], Jennifer Meyer[1]**
[1] IPN – Leibniz-Institute for Mathematics and Science Education
[2] University of Hildesheim
steinbach@leibniz-ipn.de

**ABSTRACT**: **Poster:** Automated scoring based on artificial intelligence (AI) enables more frequent, performance-contingent feedback for written texts. While learners need corrective (i.e., negative) feedback for improvement, feedback that mirrors failure may harm motivation, making it necessary to design motivating automated feedback messages especially for low-performers. To mitigate the potentially demotivating impact of explicit incorrectness cues, literature suggests providing elaborated feedback (EF) information. While EF necessarily includes information on answer correctness as it mirrors performance, the cue can be adjusted to be more implicit or explicit. This study compares the effects of more implicit and explicit incorrectness cues in AI-based automated, negative EF on student motivation and performance. After completing an English writing task, $N$ = 104 ($M_{age}$ = 13.97 years) low-performing students received EF either with or without explicit cues of incorrectness. We examined pre-post effects and group differences in motivational and performance outcomes. Data show that student performance improved but their motivation declined throughout the unit in both feedback conditions. The presence of explicit cues of incorrectness impacts the effects on self-concept, but not intrinsic value. Our findings highlight the need for further research into the design of negative feedback to ensure that it effectively motivates and supports low-performing learners.

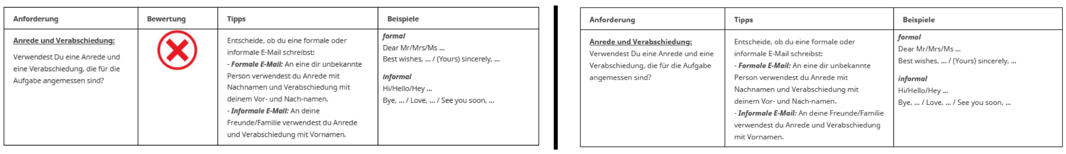
**Keywords**: Feedback, Learning Analytics, Academic Motivation, Artificial Intelligence

## 1 OBJECTIVES AND BACKGROUND

Formative assessment that measures the quality of given answers (Sadler, 1989) often involves feedback for learning purposes (Hattie & Timperley, 2007). While English writing skills are an important competence that should be fostered, they are complex and time-consuming to measure (Weigle, 2002). We addressed this issue by using an AI-based algorithm that automatically assesses students' texts in a digital learning environment, investigating the motivational effects of two different automated feedback strategies. Situated Expectancy-Value Theory (SEVT; Eccles & Wigfield, 2020), one of the theoretical frameworks central to the research of many scholars investigating the effects of feedback on motivation (Fong & Schallert, 2023), divides students' achievement motivation into self-concepts and subjective task-related values. Prior empirical evidence showed that both dimensions can be impacted by feedback (Kuklick & Lindner, 2023). *Elaborated Feedback* (EF) has been shown to benefit learning outcomes more than simpler feedback types (Mertens et al., 2022). Compared to confirmatory feedback that has been shown to motivate students by mirroring success, negative (i.e., non-confirmatory) feedback can have inverted effects (i.e., demotivating especially students with low initial performance level that receive negative feedback more often; see, e.g., Kuklick & Lindner, 2023). This demotivation may lead to insufficient effort, impairing their learning progress. However, when presenting EF, *Knowledge of Results* (KR) feedback cues that state whether a student's response was correct or incorrect are part of the feedback message (Hattie & Timperley, 2007). To lessen potential detrimental motivational effects of negative writing evaluation for low-performing learners, we manipulated the presence of an explicit KR component in the feedback message, arguing that excluding the KR cue (i.e., clearly stating “Wrong”) might benefit learners because they are confronted less explicitly with perceived task-related failure. We investigated this hypothesis in an experimental study using an AI-based algorithm to assess young writers’ performance.

## 2 METHODS

The AI-based algorithm, utilizing a neural sequence tagging model, was trained in prior studies to segment English email texts into distinct segments with quality labels, and used to score students' texts according to five writing criteria in a formal email writing task (*All information, salutation and farewell, subject, introduction and closing sentence, language style*; Horbach et al., 2022). As we were interested in low-performing students, the final sample refers to students fulfilling none of the five criteria according to the algorithm scoring and consisted of $N$ = 104 seventh- to ninth-grade students ($M_{age}$ = 13.97; $SD_{age}$ = 1.06) collected from $k$ = 48 German classrooms. Both automated feedback conditions of our randomized pre-post design provided EF information (hints and examples) on how a student could improve for each criterion. The “EF with explicit KR” feedback ($n$ = 47) provided both EF and negative KR for each criterion, whereas the “EF without explicit KR” feedback ($n$ = 57) only displayed EF regarding unfulfilled criteria without an explicit KR cue (Figure 1). Intrinsic value (4 items) and self-concept (3 items) were measured using an adapted version of a self-report instrument from Möller and Bonerad (2007; $\alpha \geq .72$.)



| Anforderung | Bewertung | Tipps | Beispiele |
|---|---|---|---|
| **Anrede und Verabschiedung:** Verwendest Du eine Anrede und eine Verabschiedung, die für die Aufgabe angemessen sind? | | Entscheide, ob du eine formale oder informale E-Mail schreibst: - ***Formale E-Mail:*** An eine dir unbekannte Person verwendest du Anrede mit Nachnamen und Verabschiedung mit deinem Vor- und Nach-namen. - ***Informale E-Mail:*** An deine Freunde/Familie verwendest du Anrede und Verabschiedung mit Vornamen. | ***formal*** Dear Mr/Mrs/Ms ... Best wishes, ... / (Yours) sincerely, ... ***informal*** Hi/Hello/Hey ... Bye, ... / Love, ... / See you soon, ... |

| Anforderung | Tipps | Beispiele |
|---|---|---|
| **Anrede und Verabschiedung:** Verwendest Du eine Anrede und eine Verabschiedung, die für die Aufgabe angemessen sind? | Entscheide, ob du eine formale oder informale E-Mail schreibst: - ***Formale E-Mail:*** An eine dir unbekannte Person verwendest du Anrede mit Nachnamen und Verabschiedung mit deinem Vor- und Nach-namen. - ***Informale E-Mail:*** An deine Freunde/Familie verwendest du Anrede und Verabschiedung mit Vornamen. | ***formal*** Dear Mr/Mrs/Ms ... Best wishes, ... / (Yours) sincerely, ... ***informal*** Hi/Hello/Hey ... Bye, ... / Love, ... / See you soon, ... |

**Figure 1: Automated EF with (left) and without (right) explicit KR cue of incorrectness (i.e., the red cross) on one example criterion (presence of salutation and farewell in the email).**

To assess knowledge acquisition, the algorithm scored the number of fulfilled criteria on (1) text revision and (2) a posttest transfer task. The procedure is shown in Figure 2.

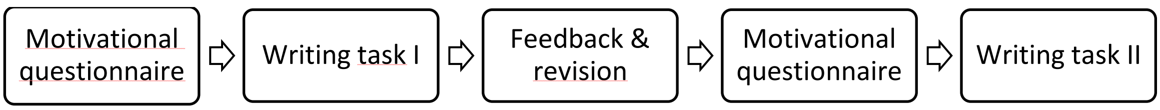



**Figure 2: Illustration of the Procedure.**

A total of four multiple regression models were employed to analyze pre-post effects and to test differences between the feedback conditions on each of the motivational and performance outcomes.

## 3 RESULTS

Students did not significantly differ in the pretest regarding their self-concept ($t$ = .468, $p$ = .641) or intrinsic value ($t$ = .285, $p$ = .776). After receiving feedback on their first text, they significantly improved performance in text revision ($t$ = 6.05, $p$ < .001, $d$ = .57) and the posttest-task ($t$ = 8.94, $p$ < .001, $d$ = .88), yet the analyses revealed a significant decrease in self-concept ($t$ = -4.39, $p$ < .001,

$d$ = -.43) and intrinsic value ($t$ = -2.68, $p$ = .007, $d$ = -.26). We found a significantly lower decrease of self-concept ($t$ = 4.39, $p$ < .001, $d$ = .24) in the “EF without explicit KR” feedback condition, but no evidence of a differential feedback effect on change of intrinsic value ($t$ = 1.51, $p$ = .131, $d$ = .09), performance improvement in the revision ($t$ = .388, $p$ = .302, $d$ = .024) or the second task ($t$ = .03, $p$ = .248 $d$ = .02).

## 4 SIGNIFICANCE

This study focused only on students with low initial performance, receiving negative performance-contingent feedback. Our results show that despite performance improvements, students’ motivation decreased, which might be a generic intervention effect as low-performers’ motivation decreased during the course of the test. The exclusion of explicit incorrectness cues in EF buffers the effect only for self-concept. Excluding KR significantly buffers detrimental effects only for self-concept, which is in line with SEVT (Eccles & Wigfield, 2020) as mirroring task-related failure is more conceptually related to self-concept compared to other outcomes (Fong & Schallert, 2023). More research is needed to understand how exactly students perceive, process, and react to negative, automated feedback. We endorse further avenues for research to take a student-centered perspective when designing AI-based feedback systems to improve motivational consequences especially for low performing students, for whom the present study demonstrated the risk of negative feedback having unfavorable effects on their self-concept and intrinsic value.

## REFERENCES


Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Fong, C. J., & Schallert, D. L. (2023). “Feedback to the future”: Advancing motivational and emotional perspectives in feedback research. *Educational Psychologist*, 1–16. https://doi.org/10.1080/00461520.2022.2134135

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Horbach, A., Laarmann-Quante, R., Liebenow, L., Jansen, T., Keller, S., Meyer, J., Zesch, T., & Fleckenstein, J. (2022). *Bringing Automatic Scoring into the Classroom – Measuring the Impact of Automated Analytic Feedback on Student Writing Performance*. 72–83. https://doi.org/10.3384/ecp190008

Kuklick, L., & Lindner, M. A. (2023). Affective-motivational effects of performance feedback in computer-based assessment: Does error message complexity matter? *Contemporary Educational Psychology*, *73*, 102146. https://doi.org/10.1016/j.cedpsych.2022.102146

Mertens, U., Finn, B., & Lindner, M. A. (2022). Effects of computer-based feedback on lower- and higher-order learning outcomes: A network meta-analysis. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000764

Möller, J., & Bonerad, E.-M. (2007). *Habitual Reading Motivation Questionnaire (FHLM)* [dataset]. https://doi.org/10.1037/t58194-000

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*, 119–144.

Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.

# Learning Analytics-based Personalized Feedback: Comparative Analysis of Behavior and Performance Trends between Groups

**JaeEun Russell, Anna-Marie Smith, Salim George, Adam Brummett**
University of Iowa
Jae-russell@uiowa.edu, anna-smith-1@uiowa.edu, salim-george@uiowa.edu, adam-brummett@uiowa.edu

**ABSTRACT**: This study compared the learning behaviors and performance trends of two groups of students in a large chemistry class: Group 1 consisted of students who exhibited low levels of engagement and performance at the beginning of the semester, and Group 2 was the remaining students. Group 1 received a series of personalized feedback tailored to their individual behavior and performance data, with the aim of fostering metacognitive reflection and enhancing engagement. Conversely, the remaining students (Group 2) received messages that were similar but generic in nature. The results revealed a significant interaction between time and group on exam performance, indicating a decrease in the performance gap between the two groups over time. Group 1 sustained their performance while Group 2's performance declined over time. However, no such interaction effect was observed in any online learning behaviors. Based on students' responses, it appears that the personalized feedback messages may have exerted a greater influence on the offline behaviors of Group 1 than on their online behaviors.

**Keywords**: Personalized feedback, Learning analytics, Higher education

## 1 INTRODUCTION

Learning analytics has been integrated into feedback practices (Banihashem et al., 2022), enabling many students to receive personalized, meaningful feedback without significantly increasing the workload for instructors (Shibani, Knight, & Shum, 2020). Several studies have demonstrated that the implementation of learning analytics facilitates the delivery of both meaningful and timely feedback to students, and have reported positive results on student satisfaction with the course and learning outcomes (e.g., Karaoglan, Yilmaz, & Yilmaz, 2022; Kohnke, Foung, & Chen, 2022). However, most of these studies have relied on students' self-reported survey and interview data to evaluate the effectiveness of learning analytics-based feedback, without thoroughly investigating whether such feedback translated into positive changes in learning behaviors. Therefore, in addition to students' self-reports and performance, this study examined the trackable learning behaviors of two groups of students over time: Group 1 who received LA-based personalized feedback, and Group 2 who received generic feedback.

### 1.1 LA-based Intervention

LA-based personalized feedback was implemented in a large chemistry course (N = 439) intended for science and pre-professional majors. At week 4 of the course semester, a group of students with low engagement was identified based on the following criteria observed during the first four weeks: 1) Attendance of less than 60% of lectures; 2) Completion of less than half of the available practice problems; 3) Earning scores below 77% on assigned homework or below 55% on the first exam. These criteria resulted in 175 students being placed in Group 1. From week 4 to week 14 of the semester, Group 1 students received a series of personalized messages with tailored information based on their engagement and performance data. These messages promoted metacognitive processing, spaced practice, interleaving, and the utilization of course resources. Furthermore, these messages were strategically designed to cultivate a growth mindset by emphasizing that students can make progress throughout the semester. Each message corresponded to the unit exams, as the instructor believed that this was when students would be most receptive to feedback about their engagement. The remaining students, referred to as Group 2 (N = 264), received three similar, but generic messages. These messages also highlighted effective study strategies and included metacognitive prompts but were of a broader nature and were not personalized based on individual student information and analytics.

# 2 RESULTS AND DISCUSSION

## 2.1 Group Behavior Trends over Time

Mixed-effect models were used to analyze changes in learning behaviors over time. The intra-class correlations for all behaviors ranged from .4 to .7 indicating between 40-70% of variation in learning behaviors was accounted for by student-level differences. Overall, the results suggest that most forms of trackable engagement decreased over time for both groups, with Group 2 exhibiting higher levels of engagement than Group 1 on average across all online behaviors.

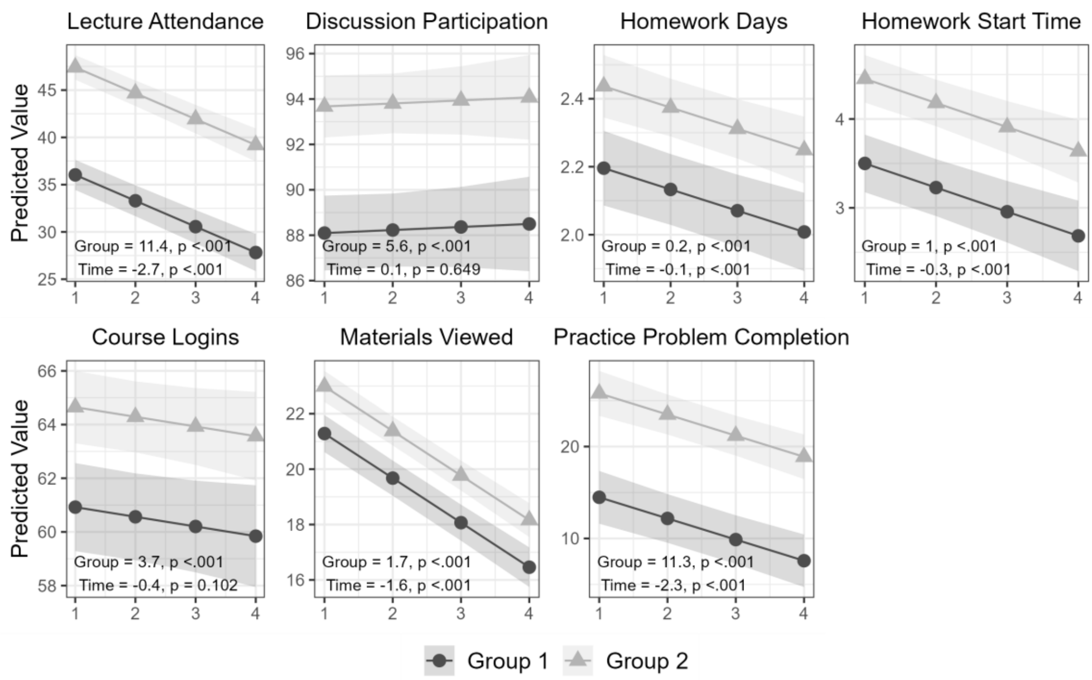



**Figure 1. Mixed effect model marginal predictions for behaviors by group. Between-group differences in engagement are consistent over time. X-axis indicates time. 1 = week 4, 2 = week 9, 3 = week 13, 4 = week 16.**

## 2.2 Group Performance Trends over Time

A mixed-effect model structure was selected to analyze changes in exam performance over time. The intra-class correlation for exam scores was .64, indicating that 64% of the variation in exam scores was accounted for by student-level differences. Results from the likelihood ratio test indicated that the addition of the Group*time interaction term significantly improved model fit, $p < .001$. The final model is shown in Table 1. There was a significant interaction between Group 1 membership and time. The performance gap in exam scores between the two groups decreased over time, as Group 1 sustained their performance while Group 2's performance declined over time.

**Table 1: Mixed effect model for exam performance**

| **Fixed Effects** | Estimate | SE | t value |
|---|---|---|---|
| Intercept | 94.3*** | 1.38 | 80.36 |
| GPA | 19.5*** | 1.40 | 13.9 |
| Group 1 | -16.4*** | 1.73 | -9.52 |
| Time | -4.1*** | 0.38 | -10.7 |
| Group*time | 4.01*** | 0.60 | 6.747 |
| **Random Effects** | $\sigma 2$ | SD | |
| Subject | 139 | 11.8 | |
| Residual | 174 | 13.2 | |

Additionally, 363 out of 439 students (Group 1 = 139, Group 2 = 224) reported their actions taken in response to these messages. Group 1 was more likely than Group 2 to take action due to feedback messages, $t(324.6) = 2.69$, $p = .007$. Specifically, 79% of students in Group 1 and 66% of students in Group 2 reported taking action due to the feedback. Furthermore, the types of actions taken by the two groups differed: Group 1 was more likely to report attending supplemental instruction and tutoring, reaching out to their teaching assistant for help, and engaging in reflective practices on their assignments compared to Group 2.

## REFERENCES

Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review*, 100489.

Karaoglan Yilmaz, F. G., & Yilmaz, R. (2022). Learning analytics intervention improves students' engagement in online learning. *Technology, Knowledge and Learning*, *27*(2), 449-460.

Kohnke, L., Foung, D., & Chen, J. (2022). Using learner analytics to explore the potential contribution of multimodal formative assessment to academic success in higher education. *Sage Open*, *12*(2), 21582440221089957.

Shibani, A., Knight, S., & Shum, S. B. (2020). Educator perspectives on learning analytics in classroom practice. *The Internet and Higher Education*, *46*, 100730.

# Mobile Learning Dropout Prediction: Identifying Key Features

**Sylwia Macinska**
Cambridge University Press & Assessment
sylwia.macinska@cambridge.org

**Lakshmi Babu Saheer**
Anglia Ruskin University
lakshmi.babu-saheer@aru.ac.uk

**ABSTRACT**: This paper investigates the dropout prediction problem specifically in mobile learning applications, with the aim to determine the best features for identifying learners at risk of disengagement. Particularly, three types of variables were compared to measure their effect on dropout prediction: session-related, performance-related and features related to self-regulated learning. A model capable of classifying learners according to their level of engagement into one of the three groups: early dropouts, high-risk learners and lows risk learners was developed, achieving an accuracy of 87%. Feature importance analyses revealed that session-related features had the most predictive power.

**Keywords**: technology-enhanced learning, student engagement, dropout prediction

## 1 INTRODUCTION

As online education grows worldwide, understanding the causes of disengagement is crucial to developing effective learning tools and practices. Despite its advantages, online learning faces higher attrition rates than traditional classrooms, influenced by factors like isolation, lack of community, and the need for strong self-regulation in learning [1, 2]. While predictive models for detecting learners at risk of dropout have been proposed in the past [3], mobile learning environments differ from traditional online courses as they tend to have limited features due to smaller screens, suffer from greater potential for distractions and differ in access to educational content. Because of this, typical dropout predictors may no longer be relevant or simply unavailable in the context of mobile learning. To address this issue, this paper investigates the dropout prediction problem specifically in mobile learning applications, using a language learning app as a case study, exploring which engagement predictors are most effective in mobile learning and whether they differ from those in traditional online platforms.

## 2 METHODOLOGY

The study utilised data from a self-study language learning app employing exercise-based instruction. Learners are presented with bundles of four activities, one for each of the language skills: reading, listening, speaking, and writing. A new bundle of activities becomes available every 24 hours, which increase in difficulty over time. The app is available to anyone, and it is free to use. The dataset included learner data from a 12-month period, considering only those who began using the app within this timeframe. The final dataset comprised 15,512 learners.

Engagement was defined based on the completion of activity bundles. Learners were categorised into three groups: (1) early dropouts (completed up to three bundles, 80% of learners), (2) high risk (completed 4 to 12 bundles, 14%), and (3) low risk (completed more than 12 bundles, 6%). These groupings were based on activity completion histograms and domain knowledge. Because of imbalanced dataset, SMOTE was applied after the train-test split to balance class representation in the training set. Three predictor categories were identified based on prior literature [3, 4] and EDA: performance-related, session-related, and self-regulated learning (SRL)-related. Performance-related features included the proportion of correct responses, response time, attempt rate, and proportion of skipped questions. Session-related features covered aspects like time to first activity, number of logins, weekend study, notification interactions, and session duration. SRL-related features, reflecting planning, monitoring, and regulating activities, included time on instructions, feedback, progress views, and learning interval.

Various machine learning (ML) techniques were considered, including Logistic Regression (LR), Support Vector Machines (SVMs), Decision Trees (DTs), Random Forest (RF), K-Nearest Neighbours (KNN), and Extreme Gradient Boosting (XGBoost). The study prioritised model interpretability, hence opaque models were not considered. The best hyperparameters for each model were identified through a Coarse-to-Fine informed search, and models were evaluated using stratified 5-fold cross-validation for generalisability. Performance accuracy was assessed on an unseen test set (a random holdout sample).

**Table 1: Final performance of each model on an unseen test set.**

| ML technique | Accuracy score |
|---|---|
| LR | 79% |
| SVM | 82% |
| KNN | 76% |
| DT | 77% |
| RF | 86% |
| XGB | 87% |

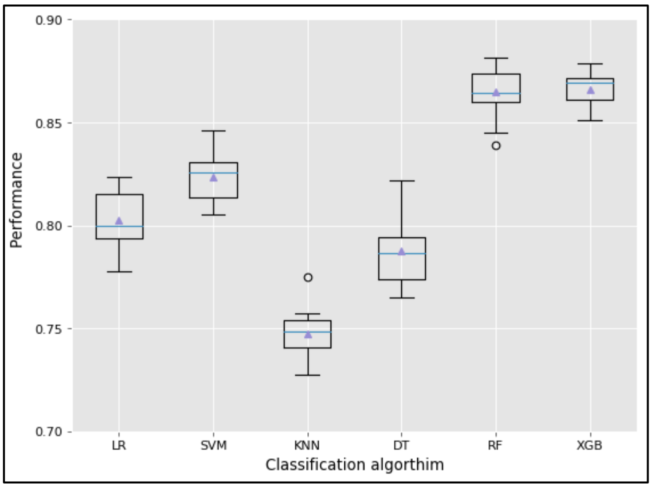



**Figure 1: Performance of ML algorithms based on 5-fold cross-validation.**

## 3 RESULTS

Figure 1 and Table 1 show models' performance as measured by accuracy score, with RF and XGBoost being the most effective. However, XGBoost showed less variability in performance and the highest performance on an unseen test set with an F1-score of 87%. Further, the model was highly accurate

(97%) in identifying early dropouts and performed well (86%) with the low-risk group. Misclassifications were more common in the high-risk group, often labelled as early dropouts.

Model interpretability is crucial in the educational domain. To understand the impact of different feature categories, the models were retrained using only performance-related, session-related, or self-regulated learning (SRL) features. Session-related features had the most significant impact (81%), followed by SRL (67%) and performance features (60%). Individual feature importance was also analysed using a drop column technique. The number of logins, learning interval, and median session time were among the most impactful features.

The study examined different time periods, ranging from 2 days to 6 weeks, to determine the shortest amount of data needed to identify early dropouts [5, 6]. The model's performance improved as more data became available. A seven-day sample period was enough to correctly classify 69% of learners, increasing to 77% after two weeks.

## 4 DISCUSSION AND CONCLUSIONS

This project aimed to predict dropout in a self-study mobile learning application, adapting features from traditional online learning research to the mobile context. Despite the absence of some features like forum activity and video-related activity, common in traditional learning platforms, the resulting model still achieved an 87% accuracy, in line with models developed for the traditional online platforms (e.g., [7]). Unsurprisingly, a trade-off was observed between the accuracy and length of the sample period. Yet, a satisfactory 77% accuracy was achieved with as little as two weeks' worth of data, with score further improved by 13%, when problem was reframed as a binary classification. Session-related features emerged as key predictors, underlining the importance of frequent interactions with the app. A model based solely on these features attained 81% accuracy, reinforcing the link between regular app usage and lower dropout risk. This aligns with traditional learning environments, where attendance is a strong engagement and performance indicator [8]. Surprisingly, SRL features, yielded only a 66% accuracy when used alone. However, the SRL indicators used in this study were rather superficial and might have not represented SRL behaviours sufficiently.

The findings of this study suggest that encouraging regular app usage can be instrumental in reducing dropout rates. This can be achieved by incorporating various features such as personalised content that aligns with learners' needs and interests, use of gamification elements like badges, leaderboards, and rewards to create a sense of competition and accomplishment, and using reminders to prompt learners to access the app.[1]

This research contributes to understanding dropout prediction in mobile learning, suggesting that even simple mobile platforms can provide sufficient data for effective dropout prediction. However, the generalisability of these findings to other apps remains to be tested, acknowledging the unique challenges and complexities of mobile learning environments compared to traditional online systems.

---

[1] Of course, engagement with a learning app, while beneficial, yields educational value only if the app itself is underpinned by sound pedagogical principles.

## REFERENCES

1. Hu, P. J. H., & Hui, W. (2012). Examining the role of learning engagement in technology-mediated learning and its effects on learning effectiveness and satisfaction. *Decision Support Systems*, *53*(4), 782-792.
2. Pardo, A., Han, F., & Ellis, R. A. (2016, April). Exploring the relation between self-regulation, online activities, and academic performance: A case study. In *Proceedings of the sixth International Conference on Learning Analytics & Knowledge* (pp. 422-429).
3. Dalipi, F., Imran, A. S., & Kastrati, Z. (2018, April). MOOC dropout prediction using machine learning techniques: Review and research challenges. In *2018 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1007-1014). IEEE.
4. Moreno-Marcos, P. M., Alario-Hoyos, C., Muñoz-Merino, P. J., & Kloos, C. D. (2018). Prediction in MOOCs: A review and future research directions. *IEEE transactions on Learning Technologies*, *12*(3), 384-401.
5. Moreno-Marcos, P. M., Muñoz-Merino, P. J., Maldonado-Mahauad, J., Pérez-Sanagustín, M., Alario-Hoyos, C., & Kloos, C. D. (2020). Temporal analysis for dropout prediction using self-regulated learning strategies in self-paced MOOCs. *Computers & Education*, *145*, 103728.
6. Xing, W., Chen, X., Stein, J., & Marcinkowski, M. (2016). Temporal predication of dropouts in MOOCs: Reaching the low hanging fruit through stacking generalization. *Computers in Human Behavior*, *58*, 119-129.
7. Liang, J., Li, C., & Zheng, L. (2016, August). Machine learning application in MOOCs: Dropout prediction. In *2016 11th International Conference on Computer Science & Education (ICCSE)* (pp. 52-57). IEEE.
8. Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59-109.

# The Future of Education: AI-Driven Data Dashboards for Effective Teaching Strategies

**Author(s): Chunsang Chan, Ada Tse, Albert Chan and Julia Chen**
The Hong Kong Polytechnic University
chun.sang.chan@polyu.edu.hk, ada.sk.tse@polyu.edu.hk, ak.chan@polyu.edu.hk, julia.chen@polyu.edu.hk

**ABSTRACT**: This poster explores the integration of Artificial Intelligence (AI) into educational data dashboards, aiming to revolutionise teaching strategies by understanding student learning patterns. The AI-integrated data dashboard, Report Toolkit (RT), comprises three innovative sub-tools: Teacher Attention Indicators (TAI), AI Insights (AII), and Icons Marking Tool (IMT). These tools are designed to enhance Learning and Teaching (L&T) practices by providing in-depth insights into student learning patterns, automating routine tasks, and facilitating strategic planning. The RT is expected to foster a data-driven culture in education, leading to more informed and effective decisions about L&T strategies. It also aims to enhance efficiency by automating the identification of anomalies and providing recommendations. Furthermore, the RT could contribute to the field of educational technology by showcasing the potential of AI and machine learning. Lastly, it could positively impact student engagement and motivation by enabling prompt identification and resolution of issues. This revolutionary approach holds the potential to transform the educational landscape, driving educational innovation and excellence.

**Keywords**: Artificial Intelligence, Education, Learning Analytics, Learning Analytics Dashboards, Learning Patterns, Teaching Strategies, Data-Driven Education

## 1 INTRODUCTION

In the rapidly evolving landscape of education, understanding student learning patterns and tailoring teaching strategies accordingly has become a paramount concern for educators worldwide. Traditional methods of data analysis often fall short in providing real-time and comprehensive insights, leading to a gap in effective teaching and learning. AI, with its ability to analyse vast amounts of data and generate actionable insights, holds the potential to bridge this gap. This document explores the integration of AI into data dashboards, a revolutionary approach that promises to transform the educational landscape. By harnessing the power of AI, we aim to equip teachers with a tool that allows them to quickly understand their students' learning patterns and adapt their teaching strategies for optimal educational outcomes.

## 2 THE NEED FOR AI IN EDUCATION

AI can significantly aid teachers by analysing learning data to provide insights that can enhance teaching methods and strategies.

AI can analyse vast amounts of data from students' performance on assignments, tests and other learning activities. This analysis can reveal patterns and trends, such as common areas of difficulty or topics that students find particularly engaging. Teachers can use this information to adjust their lesson plans and teaching methods, focusing more on areas where students struggle and leveraging topics that spark students' interest to enhance learning. During the development of LADs in the University of Queensland, Khosravi and his colleagues (2021) used a human-in-the-loop AI method to enable educators to identify, explore, and use interventions on academically at-risk students and compare their patterns to the rest of the class.

Moreover, predictive analysis using AI can forecast students' future performance based on their current learning data. This can help teachers identify students who may need additional support or intervention early on, allowing them to address potential issues before they become significant problems.

AI can also monitor students' engagement and comprehension in real-time during lessons. For instance, AI-powered systems can analyse students' responses to in-class activities or quizzes, providing immediate insights into their understanding of the lesson. This allows teachers to make on-the-spot adjustments to their instruction, such as revisiting a concept that students are finding challenging or slowing down the pace of the lesson. This real-time adaptability ensures that all students are able to follow along and grasp the material being taught, enhancing the effectiveness of the teaching process.

## 3 AI-INTEGRATED DATA DASHBOARDS

The most common application of learning analytics is the creation of online dashboards to improve learning and teaching. Numerous studies have been conducted on the concept, framework, development, implementation, and impacts of Learning Analytics Dashboards (LAD). Verbert et al. (2013) initially reviewed 15 papers on LADs. Schwendimann et al. (2016) carried out a systematic review of 55 papers on learning dashboards. Bodily & Verbert (2017) examined 94 papers related to LADs, including those on recommender systems and text messages with feedback based on LA. Over the past few years, there has been a consistent growth in the number of research studies or innovations on LADs.

To leverage the power of AI into LA, Report Toolkit (RT) will be developed. It is a groundbreaking tool that has been meticulously designed to enhance Learning and Teaching practices. It achieves this through the integration of three innovative sub-tools: Teacher Attention Indicators (TAI), AI Insights (AII), and Icons Marking Tool (IMT). Each of these sub-tools plays a crucial role in the overall functionality of the RT, contributing to its effectiveness in improving Learning and Teaching.

TAI is a sophisticated tool that leverages the power of Machine Learning to detect anomalies in data. It sifts through vast amounts of information in Learning and Teaching reports and visualisations, effectively highlighting key issues that require the attention of teachers. This feature enables teachers to quickly identify areas of concern, saving them valuable time and allowing them to focus on addressing these issues. Moreover, TAI provides valuable insights based on evidence, enabling teachers to make informed decisions and take appropriate actions.

AII, on the other hand, offers automated recommendations to teachers. This feature is particularly useful in aiding teachers in the planning, design, and implementation of effective interventions for L&T enhancement. AII is capable of providing detailed information about different student subgroups, such as performance-based, process-based, and admission-based groups. It also compares the learning process of specific groups against a control group, providing teachers with a comprehensive understanding of the learning dynamics within their classes. This information is invaluable in helping teachers tailor their teaching strategies to meet the unique needs of their students.

The third tool, IMT, is a versatile feature that allows teachers to annotate tables and visualisations with various icons. This tool facilitates note-taking, reminders, and action planning, making it easier for teachers to keep track of their thoughts and plans. Furthermore, IMT records teachers' actions, providing a comprehensive record of evidence-based L&T activities. This record can be used for future reference, helping teachers track their progress and evaluate the effectiveness of their strategies.

For instance, TAI could alert a subject leader about students not attempting online quizzes. AII could then reveal that these students also missed online classes and lecture recordings. The subject leader could use IMT to highlight these issues and plan a meeting with other teachers to address them. This scenario demonstrates how RT can effectively enhance Learning and Teaching by providing actionable insights and facilitating strategic planning.

## 4 SUMMARY

The RT is anticipated to bring about a significant transformation in Learning and Teaching by offering a data-based insight into student engagement and performance. This could pave the way for more efficient teaching methods and better student results. The RT has the potential to cultivate a data-centric approach in education, leading to more knowledgeable decisions. It could also boost efficiency by automating the detection of irregularities and offering suggestions. The RT could make a significant contribution to the educational technology sector by demonstrating the capabilities of AI and machine learning. Finally, it could have a positive effect on student engagement and motivation by facilitating quick identification and resolution of problems.

## 5 REFERENCES

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

Khosravi, H., Shabaninejad, S., Bakharia, A., Sadiq, S., Indulska, M., & Gašević, D. (2021). Intelligent Learning Analytics Dashboards: Automated Drill-Down Recommendations to Support Teacher Data Exploration. Journal of Learning Analytics, 8(3), 133–154. https://doi.org/10.18608/jla.2021.7279

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

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning Analytics Dashboard Applications. *The American Behavioral Scientist (Beverly Hills)*, 57(10), 1500–1509. https://doi.org/10.1177/0002764213479363

# OKLM: A Universal Learner Model Integrating Everyday Learning Activities with Knowledge Maps

**Kensuke Takii, Kento Koike, Izumi Horikoshi, Brendan Flanagan & Hiroaki Ogata**
Kyoto University
kensuke.takii96@gmail.com

**ABSTRACT**: This study proposes the Open Knowledge and Learner Model (OKLM), a universal learner model in which a knowledge map extracted from any domain's learning materials relates to everyday learning activities. OKLM offers various learning support, such as visualization in a dashboard, network analysis, and feedback/recommendation. To address the issue of the cost of manually extracting knowledge maps from learning materials, we present an automated method for generating them. Our experiment successfully demonstrated the generation of OKLM using this method, providing a teacher with insights into learner characteristics and structures of learning materials. Given the identified potential of OKLM, our plan includes further development as a foundational element for learning support systems.

**Keywords**: Open Knowledge and Learner Model (OKLM); Learning Analytics; learning logs; automated knowledge model composition; knowledge model visualization

## 1 INTRODUCTION

With the recent spread of online education using Learning Management Systems (LMS), a significant number of learning activity logs have been accumulated and are commonly used for Learning Analytics (LA) (Motz et al., 2019). One of the primary purposes of such online education is to expand knowledge, which requires understanding individual learners' knowledge states. Therefore, in LA research, estimating learners' knowledge states should involve the automatic connection of predefined domain knowledge models and accumulated various everyday learning activity logs.

In this study, we developed a universal learner model called OKLM (Open Knowledge and Learner Model). It automatically links any domain's graph structure of knowledge items called a knowledge map with the learning logs collected through daily learning, which promises OKLM's universality. Because of this property, the knowledge map in OKLM is "open-ended," corresponding to learning materials in any domain, and the learner model can also be "open" (Bull, 2020). This knowledge map needs to be created for each material, but we also proposed an automated way since it is costly to do this manually (Khadir et al., 2021). This method successfully generated knowledge maps from two materials, which is helpful as proof of concept to show the possibility of creating OKLM. Moreover, we developed a tool for visualizing data in OKLM, exemplifying its possibility of learning support. This study verified the effect of the tool through an experiment targeting a teacher, especially in higher education, where preparing knowledge models in advance is challenging.

## 2 OKLM FRAMEWORK

Figure 1 shows a conceptual diagram of the OKLM framework. OKLM is a learner model that can manage and track which knowledge items are covered by each learning activity by linking the learner's daily learning logs to a knowledge map generated according to the learning materials. Each node in the knowledge map stores information about the learning logs for those materials. The OKLM constructed in this manner has various possibilities to support learning. For example, it is possible to visualize a knowledge map to understand what knowledge the learner is studying during learning activities" and to acquire new knowledge through network analysis. By understanding the characteristics of learners and teaching materials, and by using statistical methods, it is possible to provide higher-order learning support such as peer learning support, material recommendation, and grade prediction.

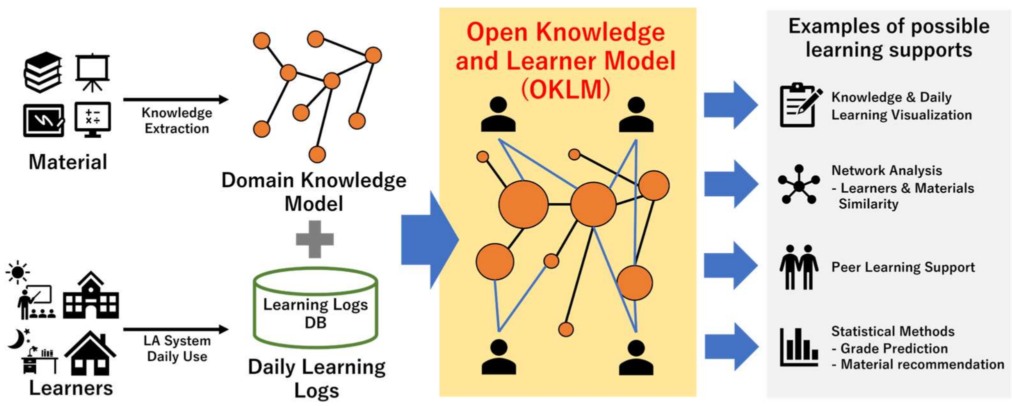



**Figure 1: OKLM Framework**

In addition, this study automatically constructed a knowledge map from learning material, considering the cost of manually constructing a knowledge map. We used the method of Flanagan et al. (2019) to create a knowledge map using the co-occurrence relations of nouns that appear on each page of the instructional material, with words as nodes and co-occurrence relations as branches.

## 3 CASE STUDY

### 3.1 Proof of Concept of Automated Knowledge Item Extraction

In this study, we automatically extracted knowledge items and generated knowledge maps using materials from an undergraduate lecture on Human Interface. Knowledge maps were generated from PowerPoint documents used in each of the two classes given by the same person. As a result, 50 nodes and 97 links were generated from the unit's material for "Various Interfaces," and 41 nodes and 74 links were generated from "Data Collection and Analysis."

### 3.2 OKLM Visualization System for Teachers

This section presents a demonstration experiment on data visualization in the OKLM, focusing on knowledge map and learning log visualization tools. The developed tool provides a visualization of knowledge maps based on user-specified subjects, students, and teaching materials. Node size in the map represents the number of learning logs associated with each knowledge item, visualized using a seek bar to adjust the date. The experiment utilized two classes and corresponding knowledge maps from section 3.1, where students used BookRoll (Flanagan & Ogata, 2018) for the e-book reader system. All learning logs from BookRoll were included in the analysis, and knowledge maps were overlaid with access logs for each class under conditions (a) and (b) (Figure 2).

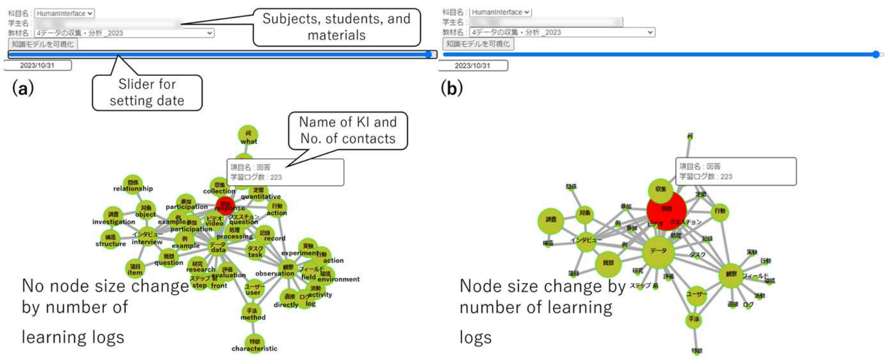



**Figure 2: Knowledge graph linked with learning logs in condition (a) and (b)**

Teacher impressions revealed that under condition (b), it was easier to obtain information on learners' knowledge and attention changes. Additionally, both conditions allowed the teacher to read the storylines, connections, and disjunctions in material content from the knowledge map structure.

## 4 DISCUSSION AND FUTURE WORK

This study proposed OKLM, a universal learner model in which knowledge maps are linked with everyday learning logs. We also showed proof of concept the possibility of creating the OKLM by an automated method for making knowledge maps from learning materials. The knowledge map constructed by this system helped understand the story of the learning material. In the future, we will improve the visualization tool for teacher support, apply it in various learning contexts, not only in the context of higher education, and promote other practical uses of OKLM. Moreover, considering the potential of OKLM to support learning and education, we intend to develop OKLM not only as a mere visualization tool but also as a foundation for various learning supports.

## ACKNOWLEDGMENTS

This work was supported by CSTI SIP Program Grant Number JPJ012347, NEDO JPNP20006, and JST Grant Number JPMJFS2123.

## REFERENCES

Bull, S. (2020). There are open learner models about!. *IEEE Transactions on Learning Technologies, 13*(2), 425-448.

Flanagan, B., Majumdar, R., Akçapınar, G., Wang, J., & Ogata, H. (2019). Knowledge map creation for modeling learning behaviors in digital learning environments. In *Companion Proceedings of the 9th International Conference on Learning Analytics and Knowledge (LAK'19)* (pp. 428-436). Society for Learning Analytics Research (SoLAR).

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

Khadir, A. C., Aliane, H., & Guessoum, A. (2021). Ontology learning: Grand tour and challenges. *Computer Science Review, 39*, 100339.

Motz, B., Quick, J., Schroeder, N., Zook, J., & Gunkel, M. (2019, March). The validity and utility of activity logs as a measure of student engagement. In *Proceedings of the 9th international conference on learning analytics & knowledge* (pp. 300-309).

# Potentials of Customizable Data-driven Dashboard: Insights from University Teacher

**Taito Kano, Changao Liang, Izumi Horikoshi, Hiroaki Ogata**
Kyoto University
kano.taito.87a@st.kyoto-u.ac.jp

**ABSTRACT**: Although various data-driven assessment dashboards have been developed, diverse learning activities in real-world contexts do not necessarily fit into their theoretical frameworks. We have developed YINSIGHT, a data-driven assessment dashboard that allows users to engineer activity indicators from trace data flexibly based on their learning activity. This dashboard contributes to data-driven assessment design in that it allows context-specific indicators with a high affinity for real-world contexts and sustainability. In this study, we conducted three interviews with a university teacher to elaborate on the applicability of YINSIGHT for assessment in her classes. As a result, the potential of YINSIGHT for assessment was demonstrated in the three perspectives of grading, improving course design, and understanding learners by customizing the indicators based on the teacher's feedback. This study moves a step towards demonstrating the specific applicability of a data-driven assessment dashboard with customizability of indicators in a real-world context.

**Keywords**: Data-driven Assessment, Customizable Dashboard, Real-world Contexts

## 1 INTRODUCTION

Many data-driven assessment dashboards have been developed to support formative assessment. However, there are still challenges in terms of sustainability in practice in real-world contexts (Gašević et al., 2022). One reason for this is that many of the activities designed in real-world contexts do not necessarily fit into the theoretical framework. Recently, the importance of customizability has been recognized in human-centered design for LADs to meet the diverse needs of users (Wise & Vytasek, 2017). We have developed YINSIGHT, a data-driven assessment dashboard that allows users to engineer indicators from trace data flexibly (Kano et al., 2023). The customizability of indicators increases the flexibility of data-driven assessment and realizes human-centered data-driven assessment which is compatible with real-world contexts. In this study, we conducted interviews with a university teacher three times to elaborate on how YINSIGHT can work as an assessment assistant for her class. In each interview, we asked her about the activities she designed, their objectives, and the potentials of the indicators engineered based on her comments. The purpose of this study is to demonstrate YINSIGHT's applicability as an assessment assistant tool in real-world context that can be achieved by its customizability.

## 2 METHOD

YINSIGHT is a data-driven assessment dashboard that allows users to engineer activity indicators freely from e-book trace data (Kano et al, 2023) as shown in Figure 1(a). The data is aggregated and filtered by context information, and transformed into raw indicators so that the means and variances are aligned. Then these raw indicators are summed based on their respective weights assigned from teacher's setting panel. The activity indicators are visualized in two panels to facilitate both a comprehensive understanding of the entire class and detailed insights into individual students. To investigate the potential use of YINSIGHT for assessment in a real-world context, we interviewed a university teacher in Japan three times as shown in Figure 1(b). The purpose is to optimize the indicators from iterations of customizing indicators based on her opinions. In each interview, after she viewed the activity indicators prepared in advance on the YINSIGHT, we asked the following questions. In the first interview, we asked the teacher (1) "Which raw indicators were important in the course activities" and (2) "What kind of activity indicators do you want to use". In the second interview, we asked (2) and (3) "What can be achieved in Grading, Improving Course Design, and Understanding Learners by using activity indicators". These categories were selected based on William's (2011) description of summative assessment and three categories of formative assessment. Lastly, in the third interview, we asked (4) "By which indicators and how their use can be realized" to elaborate on the use case. The interviewed teacher had four lectures in a course about interface design in the second semester of 2023 as shown in Figure 1(c). Each class was divided into two parts. The first half consisted of a quiz activity on the content of the previous class and group work on the assignment. In the second half, a simultaneous lecture was given using slides uploaded to an e-book. At the end of each class, a report assignment based on the class content was given. The expected activities on the e-book by students included viewing slides during class and reviewing them after class.

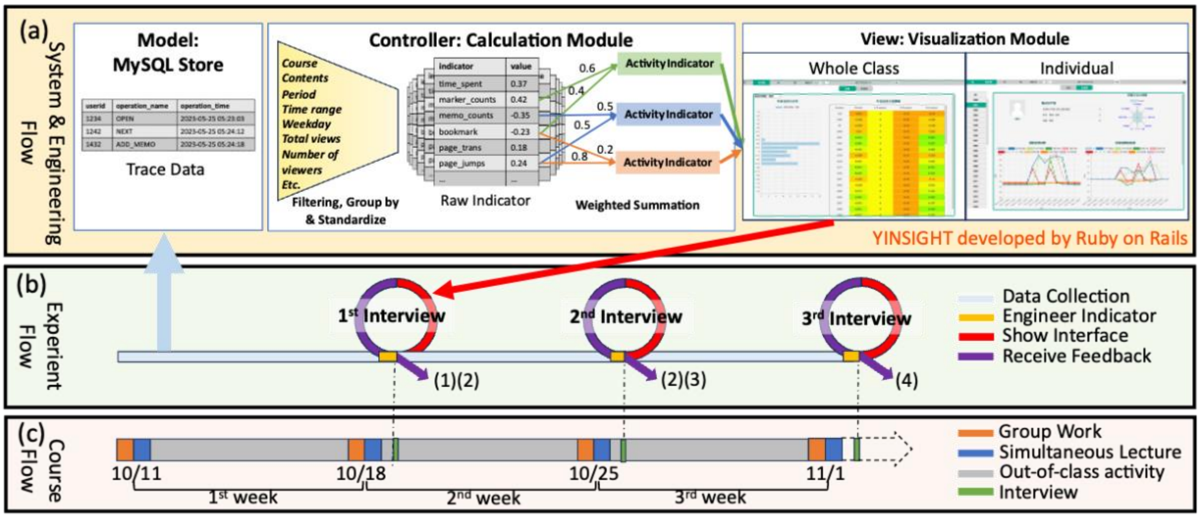



**Figure 1: Overall Flow of YINSIGHT, Experiment and Course**

## 3 RESULT & DISCUSSION

In the first interview, we displayed the activity indicators in the default settings of YINSIGHT, but from the second interview onward, the activity indicators were customized based on the prior feedback from (1) and (2). In the final interview, activity indicators such as out-of-class viewing/transitions, review viewing/transitions and night viewing/transitions were engineered based on raw indicators such as viewing time and the number of PREVs and NEXT. Table 1 shows the summary of the teacher's answers. Regarding *Grading*, her lectures were already designed to assess activities based on products. Therefore, engagement could not be an item for assessment. Thus, she pointed out the need to design activities so that activity indicators are the subject of assessment. Regarding *Improving Course Design*, the activities in her lectures were designed so that the assignments were based on the class content. Therefore, it was assumed that the review of the class slides affected the quality of the assignments and even the peer evaluation. Thus, the potential of validating peer evaluation by activity indicators was pointed out. Finally, regarding *Understanding Learners*, she was interested in different types of learning styles of high and low-performing students, such as whether they were morning or evening learners. Thus, the potential of extracting such characteristics from the relationship between activity indicators and performance scores was noted.

**Table 1: Summary of (3) What can be achieved and (4) How to realize them by indicators.**

| Objective | What (3) | Indicator (4) | How (4) |
|---|---|---|---|
| Grading | Assessing the reflection activity on the class slides for working on assignments | Review<br>Out-of-class | Design the activities in advance so that engagement in the activities is the subject of assessment. |
| Improving Course Design | Validation of peer evaluation | Review<br>Out-of-class | Correlation analysis between peer evaluation score and the activity indicators |
| Understand Learners | Comparison of learners' features between high and low performers | mid-night<br>morning<br>etc. | Correlation analysis between performance score (quiz score, peer evaluation score) and the activity indicators |

## 4 CONCLUSION

In this study, we conducted three interviews with a university teacher in Japan to investigate the potential of YINSIGHT for customizable assessment. As a result, by engineering the indicators based on feedback twice, the applicability of YINSIGHT was demonstrated in the three perspectives of *Grading*, *Improving Course Design*, and *Understanding Learners*. While this study has limitations regarding the number and type of participants, YINSIGHT can be applied to any course with daily e-book activities, which leads to further demonstration in broader contexts.


## ACKNOWLEDGEMENTS

This study is supported by NEDO JPNP20006, JSPS KAKENHI 22K20246, and 23H00505.


## REFERENCES


Kano, T., Horikoshi, I., Koike, K., & Ogata, H. (2023, December) Data-driven Competency Assessment Supporting System for Teachers. In *Proceedings of 31st International Conference on Computer in Education (ICCE 2023)*, 926-935.

Gašević, D., Greiff, S., & Shaffer, D. W. (2022). Towards strengthening links between learning analytics and assessment: Challenges and potentials of a promising new bond. *Computers in Human Behavior*, 134, 107304.

Wiliam, D. (2011). *Embedded formative assessment*. Solution tree press.

Wise, A. and Vytasek, J. (2017). Learning Analytics Implementation Design. In Lang, C., Siemens, G., Wise, A. F., and Gaevic, D., editors, *The Handbook of Learning Analytics*, pages 151–160. Society for Learning Analytics Research (SoLAR), Alberta, Canada, 1 edition.

# Classification of Answers in Math Online Tests by Visualizing Graph Similarity

**Saburo Higuchi[1,2*] and Yasuyuki Nakamura[2]**
[1]Ryukoku University, [2]Nagoya University,
[*]hig@math.ryukoku.ac.jp

**ABSTRACT**: This study explores a method to support teachers' instruction in math e-learning by classifying incorrect answers and observing how close students come to the correct answer and what misunderstandings they are likely to encounter. The proposed method first converts mathematical expression answers into labeled graphs and computes graph similarity using the tree edit distance and subtree kernel. We then drew a scatter plot of answers on a plane that reflected the distance structure with compressed information and the reduction of dimensionality with t-SNE. Finally, the plots and classifications by a human expert of classroom data were compared.

**Keywords**: math online test, classification of incorrect answers, tree edit distance, graph kernels, visualization, math e-learning

## 1 INTRODUCTION

In recent years, the rapid advancement of information technology has accelerated informatization in education, drawing increasing attention to e-learning. An essential aspect of e-learning is online testing, particularly the attention-grabbing format, which allows mathematical expressions to be used as answers (Sangwin, 2013). These formats aim to gauge the students' actual abilities by relying on the fact that students cannot answer questions if they lack an understanding of the calculation method. Therefore, an analysis of incorrect answers can reveal areas of insufficient understanding.

To achieve this, incorrect answers must be classified; however, automating this process is difficult. Although certain automated formula-scoring systems have suggested the possibility of doing so, constructing a classification mechanism is a major burden (Nakamura et al., 2021). One approach that is widely used in practice is to classify them into a small number of groups using computer algebra and pattern matching (Sangwin, 2013). The advantage of this approach is that teachers can fine-tune the classifier to reflect on their knowledge and learning objectives. However, building a classifier that appropriately addresses incorrect answers is difficult.

This study introduces a procedure for identifying the similarity among mathematical expression answers by converting them into labeled graphs and computing the graph similarity. This helps teachers classify incorrect answers and see how close students come to the correct answer and what misunderstandings they are likely to encounter.

## 2 METHODS

To collect students' answers in the form of structured mathematical expressions, we used STACK (Sangwin, 2013) questions on Moodle LMS. Because STACK grades student answers using the computer algebra system Maxima, which is built on LISP, each student's answer is automatically simplified and stored internally as a list. This list corresponds to a rooted tree graph with node labels. For example, $2/(3+x+\cos x)$ is stored as a list [/,2,[+,3,x,[cos, x]]], which is equivalent to a graph whose nodes bring labels /,2,+,3,x, cos, x, and whose edges (/,2), (+,3),. . . represent the operator-argument relationship (Fig.1).

Two approaches are used to measure the similarity between rooted labeled tree graphs. The first is the tree edit distance. Any tree can be transformed into another tree by applying a sequence of three types of operations: inserting a new node as a child of an existing node (e.g., $2/(3+x+e+\cos x)$), deleting an existing node (e.g., $2/(3+\cos x)$), and replacing the label of a node with a new label (e.g., $2\times(3+x+\cos x)$). The tree edit distance between two trees is defined as the minimum length of the sequence. An efficient algorithm (Zhang & Shasha, 1989) enables the use of this distance as a dissimilarity measure within a large collection of trees. We used the tree edit distance because the correction of their answers by the students through trial and error is directly related to tree edit operations. Another measure is the family of graph kernels (Collins & Duffy, 2001). The graph kernel function provides the inner product of the hypothetical embedding of two graphs in the vector space. We also used the distance derived from the subtree kernel, defined using the number of appearances of common subtrees. In Fig. 1, subtrees $\cos x$ and $3 + x + \cos x$ occur once, whereas $x$ occurs twice. Mathematical expressions are similar if they have many common sub-expressions.

To define meaningful classifications from an educational viewpoint, it is helpful to visualize points in a two-dimensional space. We adopted t-SNE (Van der Maaten & Hinton, 2008), which embeds them into a plane-preserving proximity relationship while discarding the details of the tree edit distances.

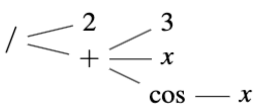



**Figure 1: Tree graph representation of $2/(3+x+\cos x)$ = [/,2,[+,3,x,[cos, x]]]**

**Table 1: Classified answers to a question $\int x^2(x-1)^5dx$ with frequency in the class**

| | expression entered | count |
|---|---|---|
| A | $(x-1)^6(21x^2+6x+1)/168+C$ | 89 |
| B | $(x-1)^6(21x^2+6x+1)/168$ | 4 |
| C | $(x-1)^6(19x^2+10x-1)/168+C$ | 4 |
| D | $(x-1)^6(7x^2-2x+2)/42+C$ | 3 |
| other incorrect answers | | 66 |
| total | | 166 |

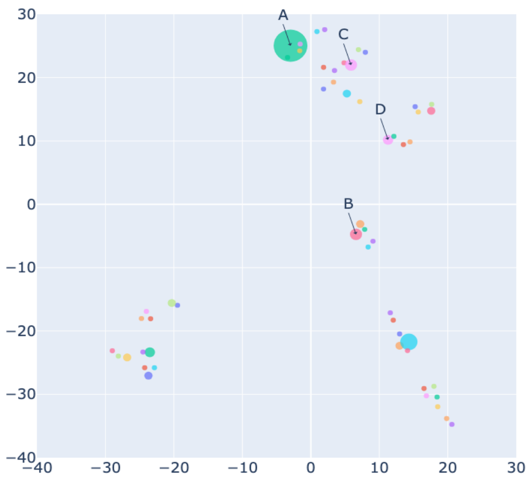



**Figure 2: Embeddings of answers into a plane by t-SNE**

## 3 RESULTS

Data were collected from an elementary university math course conducted in 2021. A total of 105 non-STEM majors participated in this course. Students were required to complete a series of online tests. Herein, we would like to report the analysis of students' answers to the indefinite integral $\int x^2(x-1)^5dx$. We converted the answers into mathematically equivalent standard forms using Maxima. The 166 submitted answers fell into 55 mathematically inequivalent categories. We calculated the distances among all trees using the Python package edist (Paaßen et al., 2015) for the tree edit distance, and the graph kernel (Sugiyama et al., 2017) for the subtree kernel. We confirmed that these two distances are strongly correlated. The embedding results for the tree edit distance are shown in Fig. 2. The size of the dots increases with frequency. A math expert with no knowledge of the current study classified the answers into groups A–D and others (Table 1). The expert gave a full grade to the correct answer A and partial grades to B, C, and D, in decreasing order. We can see that C and D are near A and are accompanied by similar incorrect answers in their neighborhoods, as shown in Fig. 2. The incorrect answer B is located farther away owing to the cost of deleting the '+' root node. We note that the embeddings depend on the random seed and do not have a preferred origin or axes in t-SNE, whereas the distances are defined unambiguously.

# 4 CONCLUSION AND DISCUSSION

We computed the distances between incorrect answers in the online math tests and visualized their similarity on a plane. This would enable one to perform clustering and automatic classification of incorrect answers and would help teachers grasp the diversity and severity of incorrect answers and perform formative assessments. Traditional grading methods assign a constant grade to 'others' that do not match any of the criteria listed in Table 1. In the current approach, as shown in Fig. 2, 'others' are decomposed into small groups. Therefore, the distance obtained using the proposed approach can automatically provide detailed partial grades. To date, the application of the proposed method has been limited to a few questions, including the examples discussed. Future research should focus on assessing the efficacy of this method by using a more extensive array of mathematical questions.

# REFERENCES

A Collins, M., & Duffy, N. (2001). Convolution Kernels for Natural Language. Advances in Neural Information Processing Systems, 14.

Nakamura, Y. et al. (2021). Automatic classification of incorrect answers to differentiation questions using potential response tree. Proceedings of Computer Aided Assessment and Feedback in Mathematics, 1113–1117. https://zenodo.org/communities/stack2021

Paaßen, B., Mokbel, B., & Hammer, B. (2015). A toolbox for adaptive sequence dissimilarity measures for intelligent tutoring systems. Proceedings of the 8th International Conference on Educational Data Mining, 632–632.

Sangwin, C. (2013). Computer Aided Assessment of Mathematics. OUP Oxford.

Sugiyama, M. et al. (2017). graphkernels: R and Python packages for graph comparison. Bioinformatics, 34(3), 530–532. https://doi.org/10.1093/bioinformatics/btx602

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. Journal of Machine Learning Research, 9(11).

Zhang, K., & Shasha, D. (1989). Simple fast algorithms for the editing distance between trees and related problems. SIAM Journal on Computing, 18(6), 1245–1262.

# ACKNOWLEDGMENTS

We thank K. Yoshitomi, Y. Miyazaki, T. Nakahara, M. Kawazoe, T. Fukui, and S. Shirai for useful discussions and critical reading of this manuscript. This study was supported by JSPS KAKENHI Grant Number JP21H04412 and a Ryukoku University Research Fellowship (SH). SH thanks Nagoya University for hospitality.

# Students Dashboard Preferences in Blended Learning with Continuous Formative Assessments

**Natalie Borter, Laura Bögli, Stefan Troche**
University of Bern
natalie.borter@unibe.ch, laura.boegli@unibe.ch, stefan.troche@unibe.ch

**ABSTRACT**: Our research explores learners' interest in a student-facing dashboard within a blended learning setting with incorporated formative assessments. Based on our experience and a literature review, we designed 20 relevant dashboard features, categorized into seven feature groups: "specific exercise feedback", "feedback on overall performance", "learning success prediction", "learning plan", "help-seeking options", "self-regulation support", and "individual learning effort". In an online survey, 265 students expressed their interest in a dashboard and chose up to five preferred features for it. More than 95% indicated a desire for a dashboard. The data analysis highlighted four feature groups that were predominantly favored by students: "specific exercise feedback", "feedback on overall performance", "learning plan" and "learning success prediction". Within the "specific exercise feedback" and "learning plan" categories, four distinct features were especially popular, each selected by more than 45% of the students. In contrast, for "feedback on overall performance" and "learning success prediction," student preferences varied regarding the type of performance comparison and time point of prediction. This study highlights considerable student interest in a dashboard designed for a blended learning setting with formative assessments, and it reveals consistent preferences for certain dashboard features. These insights are valuable for implementing dashboards aligning with student preferences.

**Keywords**: students facing dashboard, formative assessment, feedback, learning analytics, self-regulated learning, peer comparison

## 1 INTRODUCTION

Tertiary education is marked by several significant challenges: The increased number of students in classrooms, a substantial skills gap in methodological subjects, and the increasing requirement for self-regulated learning. Accordingly, individual personal feedback for each student is difficult to provide. At the same time, it is known that effective learning depends on actively engaging with the learning material and receiving individualized and continuous formative feedback (Lamotte et al., 2021). Therefore, for the past three years, the University of Bern has been offering formative assessments with immediate, automated feedback in various blended learning courses. These courses include lectures, opportunities for in-person discussions, and self-learning phases. The formative assessments, integral to these self-learning phases, provide feedback on the correctness of responses and suggestions for further learning. A prior analysis of feedback from over 1000 students revealed an interest in a learning analytics dashboard as a means to enhance their learning experience. A learning analytics dashboard is a tool to show user visual insights into learning and to improve the quality of feedback (Greller & Drachsler, 2012). However, despite their potential to enhance learning, research involving learners' needs remains very limited (Bodily & Verbert, 2017). Given the feedback from our students, we were undertaking a quantitative study to thoroughly investigate their preferences. We investigated the following hypotheses: 1) More than three-quarters of students in our learning context show an interest in a student-facing dashboard, 2) There is consistency in information students desire on their dashboard.

## 2 METHODOLOGY

A literature review (e.g. Schumacher & Ifenthaler, 2018) along with our experiences in the courses and feedback from previous students were utilized to identify and design features suitable for our learning context. A "feature" in this context, as conceptualized by Schumacher and Ifenthaler (2018), refers to a specific chunk of information presented on a dashboard. Twenty features were selected, categorized into seven feature groups: "specific exercise feedback" (three features), "feedback on overall performance" (four features), "learning success prediction" (three features), "learning plan" (two features), "help-seeking options" (three features), "self-regulation support" (three features), and "individual learning effort" (two features). Students were asked to indicate their interest in a student-facing dashboard on a scale from 0 (not at all) to 100 (very much) and choose up to five preferred features for their ideal dashboard. In our analysis, we first identified which of the seven feature groups were most chosen by students, based on their individual feature selections. Then, we examined the most popular individual features within these preferred groups. The study was approved by the ethical review committee of the Faculty of Human Sciences at the University of Bern (Nr. 2021-12-00004).

## 3 RESULTS

Among the 265 Psychology students surveyed, a substantial 96% [93% - 99%] indicated a desire for a dashboard exceeding 50 on a scale ranging from 0 (indicating no interest) to 100 (indicating a strong desire). Accordingly, far more than three-quarters of students showed an interest in a student-facing dashboard. In addition, more than half of the students selected at least one feature from the four feature groups "specific exercise feedback" (94%), "feedback on overall performance" (60%), "learning plan" (59%), and "success prediction" (55%). Table 1 shows the single features, the corresponding feature group and the selection rate for each single features selected by at least 20% of students.

**Table 1: Percentage of Selection for the most Preferred students-facing Dashboard Features**

| Feature_Group | Single_Feature | Selection_Rate |
|---|---|---|
| Specific Exercise Feedback | Links to additional formative exercises in areas where students are currently experiencing difficulties | 71% [68% - 74%] |
| Specific Exercise Feedback | Feedback on exercises (correct/incorrect) with retry options | 69% [66% - 72%] |
| Learning Plan | Learning plan showing completed and pending tasks | 56% [53% - 59%] |
| Specific Exercise Feedback | Exercises marked for repetition with retry options | 45% [41% - 48%] |
| Learning Success Prediction | Grade prediction based on prior knowledge, and module to close knowledge gaps | 30% [27% - 33%] |
| Learning Success Prediction | Grade prediction based on their current semester's performance, accompanied by improvement suggestions | 30% [27% - 33%] |
| Feedback on Overall Performance | Feedback on overall performance with comparison to own past achievement | 26% [23% - 29%] |
| Feedback on Overall Performance | Feedback on overall performance with no comparison | 20% [17% - 23%] |
| Feedback on Overall Performance | Feedback on overall performance with peer comparison | 20% [17% - 23%] |

*Notes*: *N* = 265; 95% confidence intervals in parentheses.

Four specific features from the "specific exercise feedback" and "learning plan" groups were each chosen by at least 45% of students, as shown in the first four rows of Table 1. However, for "learning success prediction" and "feedback on overall performance", there was a notable variation in preferences. No single feature in these categories was selected by more than a third of the students. Regarding "feedback on overall performance", student preferences diverged between wanting no comparison, comparisons with past performance, or peer comparison. This finding aligns with Rets et al. (2021), who observed that many students exhibit an aversion to peer comparison. In the "learning success prediction" category, students were equally divided between preferring grade prediction based on prior knowledge accompanied by modules designed to close knowledge gaps versus grade prediction based on current semester performance, coupled with improvement suggestions.

## 4 CONCLUSION

The study underscores a pronounced student interest in a student-facing dashboard, revealing a consistent inclination towards specific feature groups. Within these groups, certain individual features were favored by a majority of students, though varied preferences emerged for others. This approach highlights the significance of considering student preferences for the user-friendly and effective implementation of dashboards designed for a blended learning setting with formative assessments.

## REFERENCES

Bodily, R., & Verbert, K. (2017). Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies*, 10, 405 - 418. https://doi.org/10.1109/TLT.2017.2740172

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, 15, 42–57.

Lamotte, M., Izaute, M., & Darnon, C. (2021). Can tests improve learning in real university classrooms? *Journal of Cognitive Psychology*, 33, 974–992. https://doi.org/10.1080/20445911.2021.1956939

Rets, I., Herodotou, C., Bayer, V., Hlosta, M., & Rienties, B. (2021). Exploring critical factors of the perceived usefulness of a learning analytics dashboard for distance university students. *International Journal of Educational Technology in Higher Education*, 18, 46. https://doi.org/10.1186/s41239-021-00284-9

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.2017.06.0304

## ACKNOWLEDGEMENTS

This work was supported by a BeLEARN grant.

# A Trial Study on Understanding the Influences of Japanese Certified e-Textbook Usage in Classroom on Academic Performance Changes

**Yukako Miyanishi, Toru Nagahama, Satoshi Nakagawa, Tatsuya Horita**
Graduate School of Information Sciences, Tohoku University
Research Organization of Open Innovation and Collaboration, Ritsumeikan University
miyanishi.yukako.q8@dc.tohoku.ac.jp, toru.nagahama.a7@tohoku.ac.jp,
nak3104@fc.ritsumei.ac.jp, horita@horilab.info

**ABSTRACT**: This study explores the integration of e-books in Japanese education, specifically focusing on government-approved e-textbooks used nationwide. Investigating the relationship between e-textbook usage in English classrooms and students' English proficiency, this trial analyzes operation logs, such as the number of operations and views of in-page content. The study reveals that students with lower scores in the initial English proficiency test, who engaged more with in-page content, demonstrated improved scores in the subsequent test. This suggests a potential relationship between e-textbook usage and enhanced academic performance.

**Keywords**: e-textbook; operation log; performance change; learning log; learning styles;

## 1 INTRODUCTION

In Japan, private sector-produced textbooks for elementary and secondary education are government-certified for quality. E-textbooks are digital versions of these certified books. Despite progress in researching e-textbooks as instructor supplements, few studies explore learners' use of certified e-textbooks (Miyanishi et al., 2023). Investigating the relationship between e-textbooks usage and performance could help to improve classes throughout Japan using e-textbooks. The linkage between e-textbook operation logs and academic performance has been extensively explored in learning analytics field (Liu et al., 2022, Yildirim et al, 2022). For example, Yildirim et al. (2022) identified influential variables to final performances in their educational context. They found that the high level of learners with low-level prior knowledge may improve their performance.

The purpose of this study was to determine how students with changing performances operated e-textbooks in classroom by exploring the relationship between operation logs and performances.

## 2 METHOD

The analysis included 136 first-year junior high school students (12 years old). Scores from Eiken[1] (the English proficiency test produced by the Society for Testing English Proficiency, Inc.) administrated twice in October 2021 and February 2022 were used for the performance data. Table 1 shows the mean and standardize of scores for each group, as described later in 3.1. Operational logs in a 50-minute English class taught by one teacher (from October 2021 to March 2022) were used for e-

[1] https://www.eiken.or.jp/eiken/en/

**Table 1: Mean (*SD*) of English proficiency test scores for each group.**

| | LD (n=22) | | LU (n=15) | | MIDDLE (n=36) | | HIGH (n=44) | |
|---|---|---|---|---|---|---|---|---|
| | 1st | 2nd | 1st | 2nd | 1st | 2nd | 1st | 2nd |
| Total Score | 374.3 (73.6) | 464.6 (50.4) | 423.3 (53.6) | 389.7 (70.3) | 508.1 (29.8) | 559.9 (37.3) | 629.4 (87.6) | 681.4 (62.1) |
| Listening Score | 155.6 (58.6) | 217.2 (29.1) | 198.8 (45.5) | 170.4 (57.8) | 250 (22.7) | 264.6 (24.3) | 306.2 (49.5) | 323.8 (40.8) |
| Reading Score | 218.7 (26.1) | 247.4 (26.2) | 224.5 (16.8) | 219.3 (22.2) | 258.1 (17.7) | 295.3 (27.4) | 323.2 (46.6) | 357.6 (32.4) |

textbook usage data. The number of classes in which operations were performed, the number of operations, pages accessed, content types, and the number of times and seconds of audio/non-audio content used were tabulated for each student.

In this trial, in order to see the relationship between performances and operation logs, students were divided into 4 groups based on the score, and trends in operation logs for each group were inspected.

## 3 RESULTS AND DISCUSSION

### 3.1 Groups based on scores

The analysis included 117 students who had taken the English proficiency test both times and had performed the operation in at least one class. Figure 1 shows the results of grouping the students according to their scores: After categorizing into HIGH, MIDDLE, and LOW, the lower segment was further divided into LOW_UP (LU) and LOW_DOWN (LD) due to large variance. The categorization criteria for HIGH, MIDDLE, and LOW are as follows: Initially, 33% and 66% percentiles were calculated based on the scores in the first and second tests respectively. Next, lines l1 and l2 were determined, orthogonal to the linear regression line L and passing through the 33% and 66% percentiles. Finally, the regions of upper, middle, and lower performance were delineated using l1 and l2. LU and LD were classified based on whether they were above or below the linear regression line L.

### 3.2 Operation Logs

Of the total 341 English classes on the timetable, e-textbooks were used in 198 classes. The average number of students per class was 9.1, the average number of seconds per student was 1,534.6, and the average number of operations per student was 20.4.

Figure 2 shows the statistics of operation log by group. Students A through F correspond to the students in Figure 1; LDs may not have used e-textbooks very much, given that their values for operation was small. On the other hand, LU had slightly larger values for operation than the other

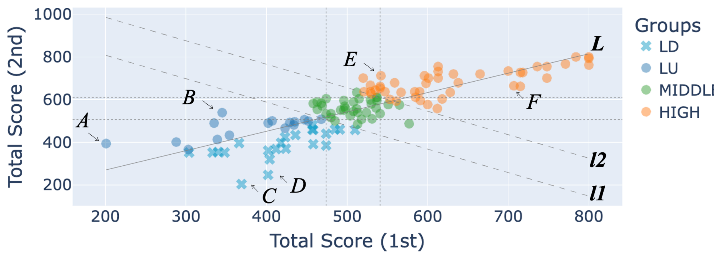



**Figure 1: Categorization for groups based on English proficiency test total scores.**

groups, suggesting that they may have been actively using the e-textbooks. The LOW group had fewer operation classes, while the MIDDLE and HIGH group had a higher number of operation classes.. As shown for students A through F, there was some correspondence between the ups and downs in scores and some operation log values in HIGH and LOW.

## 4 CONCLUSION

In this trial, from the relationship between in-class operation logs and performances, we investigated how students who had improved performances were operating e-textbooks during class. The results suggest that in the low-achieving group, the more actively students used the e-textbook, the more likely they were to increase their scores. This shows a similar feature to the results of Yildirim, D. et al, (2022). Future work is to analyze the time-series movements and to relate them to home study.

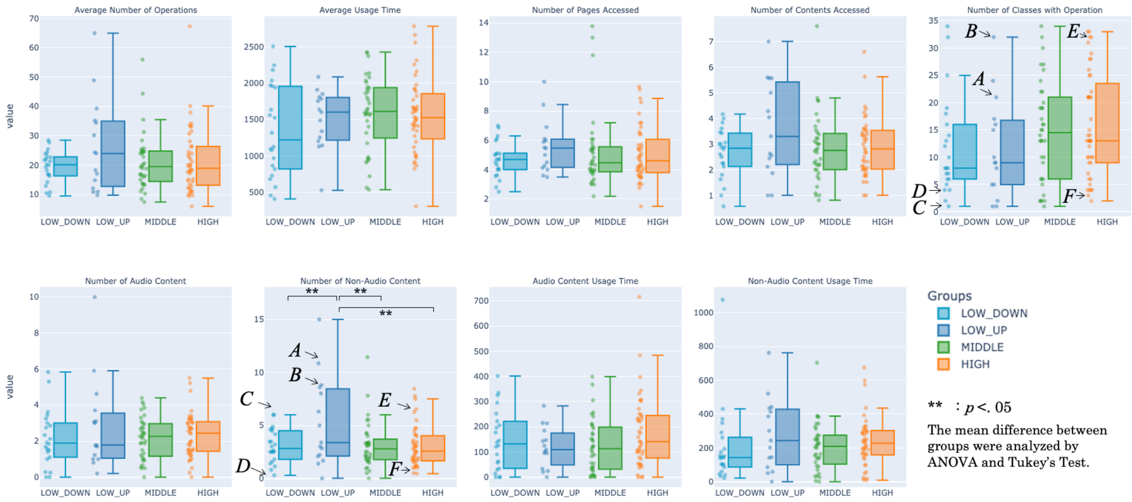



**Figure 2: Statistics of operation logs by groups based on scores.**

## ACKNOWLEDGEMENT

This work was supported by Grant-in-Aid for Challenging Exploratory Research Number JP22K18611. This study was conducted in cooperation with the Tsukuba City Board of Education, Tokyo Shoseki Co.,Ltd., Lentrance Inc. for preparing data.

## REFERENCES

Liu Y., Fan S., Xu S., Sajjanhar A., Yeom S. and Wei Y. (2023). Predicting Student Performance Using Clickstream Data and Machine Learning. Education Sciences. 2023, 13(1), 17. https://doi.org/10.3390/educsci13010017

Miyanishi, Y., Nagahama, T., Kawada, T., Horita, T., (2023). A Trial Study on the Usage of E-textbooks in Home Study at Junior High Schools for Oral Reading Assignments and Term Test of English, Japan Society for Educational Technology. 2023(2), 233-236. https://www.solaresearch.org/wp-content/uploads/2023/03/LAK23_CompanionProceedings.pdf

Yildirim, D., Gülbahar, Y. (2022). Implementation of Learning Analytics Indicators for Increasing Learners' Final Performance. Tech Know Learn 27, 479–504. https://doi.org/10.1007/s10758-021-09583-6

# Automated Classification of Student's Self-reflection Using BERT: Consideration and Possibilities for Self-Directed Learning Support

**Zixu Wang[1], Chia-Yu Hsu[1], Junya Atake[1], Izumi Horikoshi[1], Huiyong Li[2], Rwitajit Majumdar[3], Hiroaki Ogata[1]**

[1]Kyoto University, [2]Kyushu University, [3]Kumamoto University

wangzixu1999@icloud.com

**ABSTRACT:** This research aims to explore how to effectively utilize the self-reflection data generated by students during their learning process to enhance self-directed learning (SDL). In this study we collected the last 2 years' data from an SDL support environment, GOAL, with 2304 logs related to reflections, key point summaries, learning strategies, and goals statements of learners' preparation for weekly quizzes. This data can reflect students' understanding of specific knowledge points and problems and provide valuable insights for other students in the same setting. To process this data, we employed natural language processing (NLP) techniques, intending to categorize valuable contents from students' learning reflections. A BERT-based model classified contents with learning strategies and without. While the segregation model can be improved, the paper discusses the possibilities that self-reflection classifiers can play in the context of SDL support within the GOAL system.

**Keywords**: Self-Directed Learning, Large Language Models, BERT, GOAL system

## 1. INTRODUCTION

Information technology advancements are significantly transforming education. The GIGA School program in Japan, promoting digital learning with personal tablets for students, highlights the growing importance of self-directed learning (SDL). Today's learners must be proactive and self-motivated in their learning journey, identifying needs, setting goals, and evaluating outcomes (Partnership for 21st Century Skills, 2016; Knowles, 1975). In online learning environments, students generate significant textual data, including reflections and strategies. This data provides insights into their understanding and approaches to learning. We focus on utilizing these insights to enhance peer learning. Student's learning challenges and strategies can be shared to improve learning efficiency. However, these data are often underutilized. In a preliminary study, researchers tried to collect past students' challenges and solutions through questionnaires and recommend them to current students (Benedict et al.,2022). But this approach has limitations and is hard to apply widely across different courses and environments.

We aim to create a dynamic learning environment by analyzing students' textual data to promote peer learning and enhance learning outcomes. The challenge, however, lies in filtering out irrelevant information from these textual data. Hence, the research question is: How can we apply Natural

Language Processing (NLP) technology to analyze self-reflection artifacts, thereby extracting potentially valuable content?

## 2. METHOD

To meet our research goals, we used the GOAL system (Majumdar et al., 2018), which allows students to track and analyze their learning activities in real time. They can compare their performance with classmates and their historical data to assess and analyze learning outcomes and set goals. Figure 1 shows the learners' activity flow and the interface of the GOAL system in the self-analysis step.

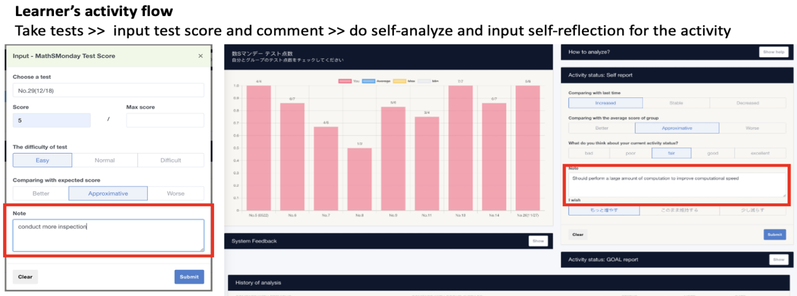



**Figure 1: Learners' activity flow and GOAL self-analysis Interface**

The dataset includes 706 students' records from April 2021 to March 2023, totaling 10,613 entries, with 2,304 entries containing self-reflections on their learning. Prior to the feature extraction and training of the text, the study conducted necessary data preprocessing. The text data were manually labeled into two categories: self-reflection notes with learning strategies (1390 entries), which include constructive feedback, discussions of challenges, and error-handling methods; and those considered without learning strategies (914 entries), characterized by a negative tone and inclusion of irrelevant content regarding the student's learning process. During the labeling process, two authors labelled independently a randomly selected portion of the data. The Krippendorff's alpha coefficient calculated for the results of the two labelers is 0.860. In this context, the notes which contain completely nonsensical text or simple complaints will be removed because they could potentially cause harm to the recommended individuals. Subsequently, to effectively tokenize the text and convert it into IDs that the model could process, we employed the BERT model specifically designed for Japanese text, cl-tohoku/bert-japanese[1].

## 3. INITIAL MODEL TRAINING RESULTS AND CONCLUSION

On average, the model achieves an average accuracy of 75.61% and the F1 Score is averaging at 79.89% in five-fold cross-validation, with each round comprising four training epochs. Our model showed improved learning from the dataset but indicated potential overfitting, as evidenced by increased validation loss. This is especially important for our study because we don't just want to analyze the current 2304 data entries; we want to build a system that continuously collects text data

---

[1] https://huggingface.co/cl-tohoku/bert-base-japanese

from students' activities and automatically recommends it to students who might need it. Figure2 shows the analysis flow and examples of text, it also shows the performance of the model.

**Analysis flow**

10613 activities entries >> 2304 self-reflection text data >> labeling >> word segmentation and tokenization >> build BERT model

| テキスト（日本語） | Text（English） | With Strategy |
|---|---|---|
| 図に書き込んでけ！ | Write it in the diagram! | ○ |
| テスト範囲をしっかり確認する | Make sure to check the scope of the test thoroughly. | ○ |
| 普段あまり使わない単語を中心に間違えてしまったので，しっかり復習しておきたい。 | I want to review thoroughly because I made mistakes mainly on words I don't usually use. | ○ |
| ひどい点数 | Terrible score. | × |
| いろいろ悲しい | Various sadness. | × |

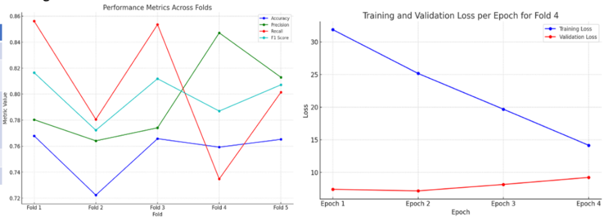



**Figure2: Analysis flow and examples of text and model performance**

## 4. FUTURE WORK

In this study, we propose an approach to classify students' self-reflective text data to aid their self-directed learning. We also defined preliminarily criteria for assessing the relevance of students' reflection data and attempted to train a model using the BERT pre-trained model for automatic classification of the text with learning strategies or not. We aim to establish a platform for information exchange and sharing, enabling students to understand the situations of their peers, thereby enhancing their learning efficiency, avoiding the same mistakes, and better cultivating their ability for self-directed learning. So, we do not insist that those who receive recommendations must follow our suggestions. For our future work, we first plan to refine the criteria for text labeling to reduce ambiguity. Next, we'll employ regularization techniques such as dropout or weight decay and consider trying different large language models (LLMs) to improve performance. Following this, we aim to analyze student's types based on their activity logs and results. Building on the results of this study, we plan to move towards an automated way to match and share relevant reflection notes as a peer support similar to the approach proposed by Benedict et al. (2022). The module can be integrated into the GOAL system to personalize and recommend potentially helpful information to learners.

## 5. ACKNOWLEDGEMENT

This work is partially funded by JSPS KAKENHI JP22H03902.

## REFERENCES

Partnership for 21st Century Skills. (2016). A state leader's action guide to 21st century skills: A new vision for education. Tucson, AZ: *Partnership for 21st Century Skills.*

Knowles, M. S. (1975). *Self-directed learning: A guide for learners and teachers*. New York: Association Press.

Majumdar, R., Yang, Y., Li, H., Akçapınar, G., Flanagan, B., & Ogata, H. (2018). GOAL: A system to support learner's acquisition of self-direction skills. *In Proceedings of the 26th International Conference on Computers in Education* (pp. 406-415).

Benedict, A., Al-Hossami, E., Dorodchi, M., Benedict, A., & Wiktor, S. (2022). Pilot Recommender System Enabling Students to Indirectly Help Each Other and Foster Belonging Through Reflections. In *LAK22: 12th International Learning Analytics and Knowledge Conference* (pp. 521-527).

# Development of a Learning Environment Supporting Self-Regulated Learning with Self-Evaluation

**Tatsuomi KIKAWA[a], Masafumi YAMASAKI[b], Kenta MIYAMOTO[b] and Takahiko MENDORI[b]**

[a]*Graduate School of Engineering, Kochi University of Technology, Japan*
[b]*School of Information, Kochi University of Technology, Japan*
265095k@gs.kochi-tech.ac.jp, {240377m, 240372r}@ugs.kochi-tech.ac.jp,
mendori.takahiko@kochi-tech.ac.jp

**ABSTRACT**: In this study, we focused on self-evaluation to support self-regulated learning. While it is understood that learners' self-evaluation contributes to the improvement of task achievement and self-regulation abilities, it is challenging for instructors alone to have learners input self-evaluations and reflect them in the class. We developed a system that allows learners to input self-evaluations for each page of learning materials and developed a system that aggregates these self-evaluations and displays them to instructors. This was done to support learners' self-regulated learning and incorporate their self-evaluations into the class. We had learners use this system in actual classes and collected learning logs. It became evident that improvements are needed in the self-evaluation input process.

**Keywords**: self-regulated learning, self-evaluation, feedback

## 1 INTRODUCTION AND RELATED WORK

Today, in Japanese education, emphasis is placed on active learning and individually optimized learning, where learners proactively engage. This approach aligns closely with self-regulated learning. This study aims to promote active learning among learner by providing support for self-regulated learning. Self-regulated learning involves learners actively engaging in their learning process in terms of metacognition, motivation, and behavior. Specifically, we are guided by a social cognitive model consisting of three cyclic stages: anticipation, performance, and reflection (Zimmerman, 1989, 2011). Research on supporting self-regulated learning often involves implementing courses with self-regulated learning, which can be burdensome for educators. To address this, there is consideration of utilizing systems to provide similar support. In related research, Khiat et al. have reported on a system that supports self-regulated learning, including prompting self-evaluation regarding tasks and learning materials (Khiat, 2022). Self-evaluation, a component of the self-regulated learning process, plays a crucial role in motivation. Research on self-evaluation suggests that frequent positive self-evaluation enhances self-efficacy, leading to improved task accomplishment and self-regulation abilities (Schunk & Ertmer, 1999). Subsequent studies emphasize the importance of accuracy in self-efficacy and self-evaluation, indicating that appropriate self-efficacy and self-evaluation can be achieved through self-reflection (Zimmerman, 2011). It is challenging for a single instructor to regularly prompt self-evaluation and provide feedback. As mentioned above, we're developing a system that allows even a single instructor to understand and leverage learner' self-evaluations for exercises. We aim to enable learner to input self-evaluations, specifically for each page of learning materials, enhancing the accuracy of the self-evaluations. Furthermore, by aggregating and sharing these self-evaluations, an instructor can comprehend learner' self-evaluations status and utilize it for exercises. As an additional feature, we will incorporate the display of learning times as an indicator for self-evaluations and provide recommendations for learning materials based on self-evaluations. We extend our system allow learners to enter self-evaluations for each page of the learning material (Mori, 2019).

# 2 SYSTEM OVERVIEW

## 2.1 Expansion of Our System for Self-Evaluation

The extended system is a PDF viewing web application designed to collect learning behavior, such as page-by-page viewing of learning materials. It comes equipped with standard features like page jumps, scaling adjustments, and annotations. As an extension of this study, a form has been added below the existing system, allowing users to input self-evaluations corresponding to the learning materials, as illustrated in Figure 1.a. The self-evaluation input field switches dynamically with each page change. Upon the first visit, the default entry is set to "-", and users can input a five-level self-evaluation (o: understand, !: important, -: usually, ?: inquire, x: incomprehensible). By inputting self-evaluations during class, learner can grasp their understanding in real-time. This behavior leads to a clear identification of learning materials or pages to review during study, contributing to motivation and resulting in a sense of satisfaction through the learning process.

## 2.2 Self-Regulated Learning Support System

The developed system is designed to display learners' self-evaluations and viewing times of learning materials. Figure 1.a shows an interface for inputting learners' self-evaluation, which is an expansion of our previous system (Mori, 2019). It is divided into functionalities for learner and instructors/administrators. For learner, as showed in Figure 1.b, features include recommending learning materials, aggregating self-evaluations, visualizing class progress, and summarizing learning times. Each page transition visualizes the viewing time and self-evaluation for each session. This clarity helps learner identify areas to focus on during previewing or reviewing, contributing to motivation.

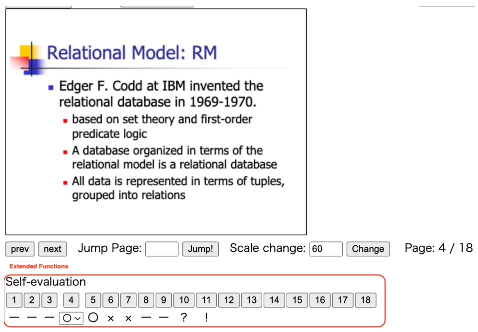
# 3 EXPERIMENT

## 3.1 Experiment Overview

We conducted an experiment to evaluate our system. The targeted course was the Discrete Mathematics (117 students enrolled) at our university, held from October 3rd to November 28th. Throughout this course, which consisted of 14 sessions lasting 90 minutes each, the system was made available consistently, and the usage history and learning records were collected.

## 3.2 Result

The learning records indicate that 80 learners used the developed system at least once, and 31 learners voluntarily entered self-evaluations. The number of times learner viewed learning materials through recommendations was 15, and there were 6 learner who used self-evaluations frequently (showing traces of entering self-evaluations over 100 times). However, the survey results revealed



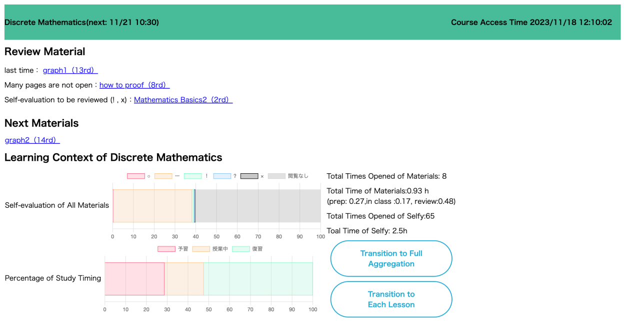




**Figure 1.a. Expansion of Mori's System** **1. b. Developed System Experiment**

that approximately 80% of learner who responded to the survey hardly used the system. The primary reason cited was that it seemed "troublesome," accounting for 70%. Additionally, while over 90% of learner indicated that previewing and reviewing were either essential or dependent on the content, 75% admitted to occasionally or not engaging in these practices. Therefore, it is believed that some form of support is needed. The gap between the survey and learning logs is attributed to the presence of trial usage records.

## 4 CONCLUSION

In this study, we focused on self-evaluation to support self-regulated learning, developing a system that allows learners to input self-evaluations for each learning material. Subsequently, we had learners use this system in actual classes and gathered feedback through surveys and learning logs. While the survey results indicated the need for some form of support for previewing and reviewing, further consideration is required to determine how to effectively utilize the system for this purpose.

## REFERENCES

Khiat, H., & Vogel, S. (2022). A self-regulated learning management system: Enhancing performance, motivation and reflection in learning. *Journal of University Teaching & Learning Practice,* 19(2), 43-59.

Ministry of Education, Culture, Sports, Science and Technology (2023). Integrated enhancement of "individualized optimal learning" and "collaborative learning." Retrieved Oct. 22, 2023, from https://www.mext.go.jp/a_menu/shotou/newcs/senseiouen/mext_01317.html (in Japanese)

Mori, Y., Sakamoto K. & Mendori, T. (2019). Development of a Real Time Viewing Status Feedback System and its impact, *Companion Proceedings 9th International Conference on Learning Analytics & Knowledge (LAK19),* 2019.3.

Schunk, D. H., & Ertmer, P. A. (1999). Self-regulatory processes during computer skill acquisition: Goal and self-evaluative influences. *Journal of Educational Psychology,* 91, 251-260.

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning, *Journal of Educational Psychology,* 81, 329-339.

Zimmerman, B. J., Moylan, A., Hudesman, J., White, N., & Flugman, B. (2011). Enhancing self-reflection and mathematics achievement of at-risk urban technical college learner, *Psychological Test and Assessment Modeling,* 53(1), 141-160.

# CAPP Logger: Proof of Concept of a New Writing Analytics Methodological Approach

**Yoram M Kalman**
The Open University of Israel
yoramka@openu.ac.il

**ABSTRACT**: This Poster describes proof of concept work on a new methodology for logging and for performing Writing Analytics. This approach uses a logger that contemporaneously records both the writing process and the evolving textual product of that process. The poster describes how large-scale automated analysis of the logger's output allows novel Writing Analytics insights.

**Keywords**: writing analytics, CAPP Logger, temporality, editing

## 1 INTRODUCTION

Writing Analytics (WA) focus on using quantitative tools to derive insights about the ways students and other users produce written texts (Palmquist, 2019). Most WA focuses on either analyzing the *product* of the writing – the text that the user produced, or on the *process* through which the user created the text – the keyboarding and pointer device actions of the user. CAPP Logger is a tool that unifies these two methodological approaches. It records the keystrokes and mouse clicks of the user, as well as the evolving text the user is producing, while keeping the logging of all of these activities fully synchronized and saved in one log. In this poster we demonstrate how analyzing 552 such logs of sessions in which undergraduate students created short descriptive texts can lead to new insights about the way users create and edit texts.

## 2 METHOD

CAPP Logger (Contemporaneous Analysis of Process and Product Logger) is based on a web-based interface where participants in the experiment were instructed, after an informed consent page, to type a description of a vacation they took or an event they participated in (e.g. party, hike, cultural event, family event). The participants were asked to type for at least seven minutes, to describe their experience in detail, and to pay attention to clarity and ease of reading. The participants were undergraduate students in a management course who participated in a series of online experiments in return for course credit.

CAPP Logger logs the time-stamped key- and mouse-strokes as well as the text as it evolves, and the log is saved as a json file. This file is analyzed using Python code that applies a taxonomy in order to identify the "history" of each of the (1) words and (2) sentences that were typed by the users including the creation of new words and sentences, their editing and modifications, and their deletion. These are used to create a host of descriptive and analytic variables. These include count variables such as number of words created, deleted or edited, temporal variables such as average time for typing a word or a sentence, and longitudinal variables that measure count and temporal variables as they evolve along a writing session. These longitudinal variables were calculated by dividing the writing session into ten equal periods of time, and assigning a number for each decile, resulting in variables such as new words per decile (count) or median time per word (temporal). A limited sample of these variables were selected for the proof of concept analyses presented below, with the goal of evaluating the potential of this approach to provide new insights in WA research. Widely used data preparation, processing and analysis methods were employed, and minimal technical and methodological details are presented here due to space limitations of the poster format.

## 3 RESULTS

Texts produced by 552 different participants were analyzed. Table 1 details some of the descriptive statistics of the writing sessions:

**Table 1: Descriptive statistics of the 552 analyzed writing sessions.**

| | # of words | # of words incl. deleted words | # of sentences | # of sentences incl. deleted sentences | avg. time typing a word | median time typing a word | avg. time typing a sentence | median time typing a sentence | word edits | sentence edits |
|---|---|---|---|---|---|---|---|---|---|---|
| Mean | 142.7 | 161.0 | 10.2 | 12.6 | 2.6 | 0.9 | 25.0 | 13.5 | 9.5 | 9.4 |
| Std.dev | 58.2 | 62.9 | 4.7 | 5.7 | 2.2 | 0.3 | 16.3 | 9.2 | 7.4 | 7.3 |
| Median | 140.0 | 157.0 | 9.5 | 12.0 | 1.8 | 0.8 | 20.8 | 11.1 | 8.0 | 8.0 |
| Skew | 0.3 | 0.3 | 0.7 | 0.7 | 1.8 | 1.0 | 1.7 | 1.3 | 1.0 | 1.0 |
| Kurtosis | 0.3 | 0.1 | 0.6 | 0.6 | 2.8 | 1.8 | 3.7 | 1.8 | 0.7 | 0.8 |

Figure 1 presents graphs of four longitudinal variables that plot the average for all 552 participants of each of the four variables through the writing session. The variables were min-max scaled and proportionally scaled.

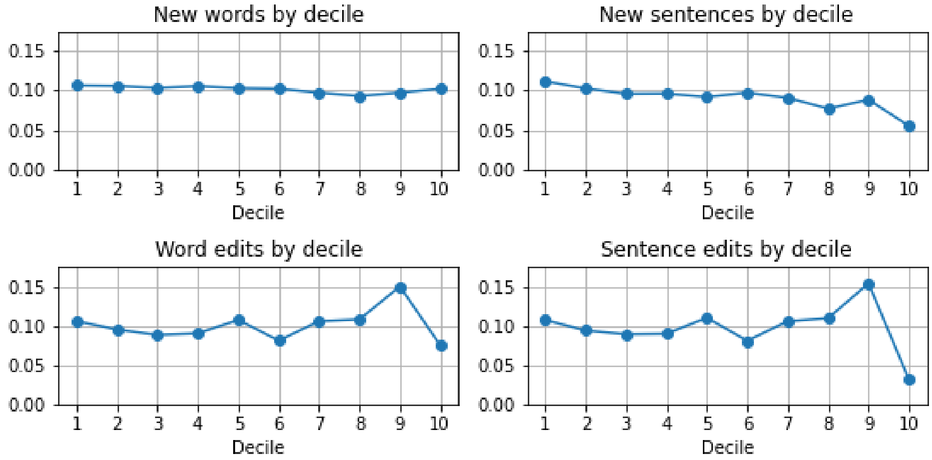



**Figure 1: An example of four longitudinal count variables and their distribution through the ten deciles of the writing sessions of all users (average)**

The final analysis that is presented here is a regression analysis that looked at 13 descriptive statistics of the writing sessions, and explored their relationship to the skewness of the distribution of the "new words by decile". As can be seen in the top left graph in Figure 1, the overall skewness of the distribution of new words is low and the line is relatively flat. But the hundreds of distributions that comprise this average skewness (-0.19) have a wide range of skewness levels (-2.44 to 3.16), and the regression allows an exploration of how that skewness is related to the 13 descriptive variables. After correcting for collinearity, four variables were removed and the following results were calculated:

```
Mean Squared Error: 0.6330935223426764
                                       Coefficient   Probability
words count                              -1.144028   0.399384
sentences count                          -0.533575   0.670137
average time for typing a word           -0.311455   0.722687
median time for typing a word            -0.926639   0.254434
average time for typing a sentence       -0.627889   0.537449
median time for typing a sentence         0.172837   0.793739
median time for starting new sentence    -0.024860   0.969570
sentence edits                            0.018724   0.975670
average words in a sentence              -0.484020   0.693359
R-squared: 0.11697092533722642
```

## 4 DISCUSSION

This poster reports preliminary results from CAPP Logger, a novel approach to WA that logs and analyzes both the writing process and the product of that process as the writing continues. Table 1 describes a sample of the wide range of variables that can be extracted using CAPP Logger, including variables that are difficult to attain unless we closely track longitudinally the evolution of each word and sentence as they are written, edited and even deleted. Figure 1 describes the average distribution of some of the variables throughout writing sessions, demonstrating a way to profile the longitudinal activity of different users. We can see, for example, that on average the rate at which new words are produced by participants is relatively even, and that the number of word and sentence edits slightly peak and then dip around the middle of writing sessions, and drop towards the end of the sessions. The regression that explores the diversity behind the relatively even average rate of creating new words shows, for example, that variables such as a higher total word count and with longer median time for typing words might be associated with a more negatively skewed rate of producing new words, while longer median time for typing sentences and a higher number of sentence edits might be associated with a more positively skewed rate of producing new words. The high probabilities of the regression coefficients and the low R-squared of the regression suggest that these preliminary findings should be treated as very tentative, yet the findings demonstrate the potential of CAPP Logger to create more nuanced and detailed characterizations of writing and editing activities, as well as to create input for LLM-supported analysis of the Logger's logs, and to analyze AI-aided writing.

## 5 ACKNOWLEDGEMENTS

I wish to thank Laura Allen for helpful discussions of CAPP Logger and of WA taxonomy, and Yuval Erez, Sveta Gimpelson and Itay Zandbank for their thoughtful and creative technical support.

## 6 REFERENCE

Palmquist, M. (2019). Directions in Writing Analytics: Some Suggestions. *The Journal of Writing Analytics*, *3*(1), 1–12. https://doi.org/10.37514/JWA-J.2019.3.1.02

# Writing Analytics for Assessing Students' Decision-Making in Computer-Based Open-ended Situational Judgment Tests

**Jinnie Shin**
University of Florida
Jinnie.shin@coe.ufl.edu

**Eric Chen**
University of Florida
Chen.Eric@ufl.edu

**ABSTRACT**: Computer-based open-ended Situational Judgment Tests serve as an effective means of evaluating individuals' problem-solving and ethical decision-making skills. However, scoring these tests presents a challenge due to the inherently unstructured nature of the responses, which often include essay-style writing and lack clear-cut solutions. We employ Transformers, to systematically decode the intricate and unstructured responses in open-ended SJTs. We examine and compare the performance of several transformer algorithms. Our results indicate that these pre-trained models achieve close to a .90 recall rate in accurately understanding both the claim and the position of the argument.

**Keywords**: situational judgment test, transformer models, argument mining, writing analytics

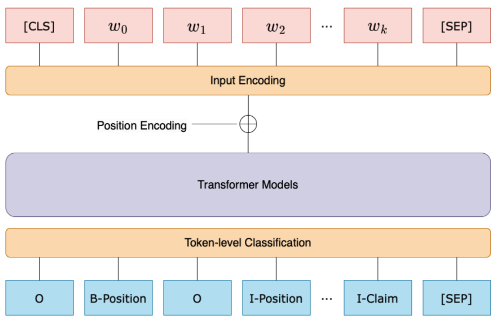
## 1 INTRODUCTION

Computer-based Situational Judgment Tests (SJTs) are increasingly important in healthcare education for assessing complex skills like problem-solving and ethical decision-making (Patterson et al., 2016). These tests, using multimedia and textual prompts, evaluate crucial non-cognitive attributes such as professionalism and teamwork by presenting real-life scenarios. However, their open-ended nature, often involving morally ambiguous situations, makes scoring prone to biases and the feedback process laborious (Schmitt et al., 2019). To address these challenges, this study employs deep neural language models, specifically Transformers, for systematic analysis of SJTs' narrative responses. Our research, conducted in two experiments, aims to identify patterns in student arguments and investigate potential scoring biases. The first experiment focuses on classifying argument elements using transfer learning, and the second examines themes within these arguments, particularly considering the impact of students' diverse educational backgrounds on their argument quality.

## 2 LITERATURE REVIEW

The ability to effectively construct and discern arguments in narrative writing is crucial in educational contexts, as it reflects critical thinking skills and rhetorical proficiency. This is particularly relevant in computerized SJTs, which assess these competencies through narrative components. These tests, benefiting from adaptability and scalability, require sophisticated methods for response analysis to ensure accurate evaluation of argumentative elements. This precise evaluation is essential for the fairness and effectiveness of the tests, especially in high-stakes academic or professional environments. Moreover, in advanced academic settings, the way students develop, present, and justify their arguments significantly impacts their academic success and future employability prospects. The critical role of argument understanding in academic settings necessitates advanced techniques for analyzing narrative writings, where Named Entity Recognition (NER) methods, particularly transformer-based models, have shown promise (Galitsky et al., 2018). These models excel in diverse NLP tasks, including argument element identification, as evidenced by applications like the Argument Reasoning Comprehension Task (ARCT). Recent research has utilized these models for various argument identification tasks, such as in the AMCT corpus and interactive argument mining. However, a major challenge lies in the need for extensive, accurately annotated data. Transfer learning offers a solution, allowing models trained on large datasets to be fine-tuned on specific, smaller datasets, thereby mitigating the data scarcity issue. This approach has shown potential in improving model performance on target tasks, even with different label spaces or network architectures. Nevertheless, further exploration is required to optimize transfer learning strategies and fine-tuning methods for complex argument detection tasks (Yamaguchi et al., 2022).

## 3 METHODS




In our study, we analyzed narrative responses from 878 students across the United States, Canada, the United Kingdom, and Australia, who participated in a computerized, video-based situational judgment test (Dore et al., 2017). This test generated 10,536 responses to 12 scenarios, where students expressed their views supported by logical claims and evidence. The demographic profile included 63% Bachelor's degree holders, 16% with advanced degrees, and a significant portion from urban and larger town settings. The majority were in the 20-25 age group, with a gender distribution of 403 males, 874 females, and 19 non-binary individuals. On average, responses comprised approximately 215 tokens across 9 sentences. Additionally, we utilized a secondary dataset from Kaggle's Feedback Prize competition, consisting of 15,594 argumentative essays by U.S. students, annotated for key argument components like 'position' and 'claim'.

## 4 RESULTS

Tables 1 and Figure 2 provide a detailed assessment of the transformer model's efficacy across both datasets. Table 1 depicts the accuracy rates of the transformer models for the SJT responses. To enhance the robustness of our findings, particularly within the realm of narrative writing, we engaged two human evaluators to scrutinize the SJT response dataset results. Utilizing the weighted-box fusion technique, the models proficiently detected the "position" and "claim" elements in the narrative writing samples. The reliability between the two raters, by their exact agreement, resulted in scores of 0.80 and 0.73 for the "position" and "claim" segments, respectively.

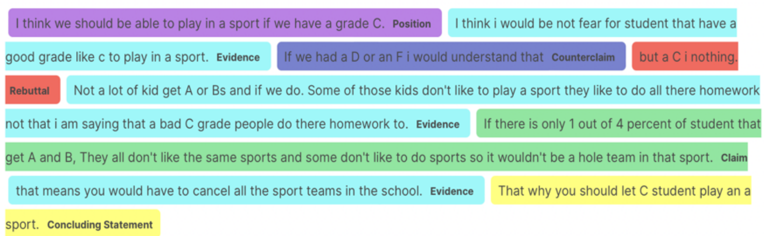



**Figure 2: Example Narrative Writing with the Argument Elements Identified**

**Table 1: Classification Performance Results of the Transformer Models and Weighted Box Fusion**

| | Claim | | Position | |
|---|---|---|---|---|
| Model | Confidence | Recall | Confidence | Recall |
| bigbird-roberta-base | 0.626 | - | 0.68 | - |
| funnel-large | 0.617 | - | 0.72 | - |
| deberta-large | 0.642 | - | 0.717 | - |
| deberta-large-v2 | 0.624 | - | 0.726 | - |
| deberta-lstm-jaccard | 0.602 | - | 0.732 | - |
| deberta-xlarge | 0.642 | - | 0.727 | - |
| longformer-large | 0.601 | - | 0.693 | - |
| facebook-yoso | 0.62 | - | 0.691 | - |
| Weighted box fusion | 0.478 | 0.832 | 0.638 | 0.887 |

## 5 SIGNIFICANCE OF THE STUDY

The significance of this study lies in addressing the intricate challenges presented by computer-based open-ended SJTs, which are pivotal in evaluating critical non-cognitive skills like decision-making and ethical reasoning in educational settings (Dore et al., 2017). By deploying advanced writing analytics, specifically the use of pre-trained deep language models such as Transformers, this research marks a substantial advancement in the systematic analysis and scoring of SJTs. The study's methodology, which combines the nuanced capabilities of transformer models with unsupervised keyword extraction, stands to contribute significantly to the field of educational assessment by providing a more objective and reliable measure of students' decision-making capabilities. This is particularly relevant in high-stakes environments where the quality of argumentation directly influences academic and professional outcomes. By demonstrating that these models can achieve a high recall rate in discerning the claim and position within arguments, the research opens new avenues for fair and efficient assessment practices. This has profound implications for the educational landscape, potentially transforming how students' critical thinking and problem-solving skills are nurtured and evaluated, ensuring that learners from diverse backgrounds are assessed with equity and precision.

## REFERENCES

Dore, K. L., Reiter, H. I., Kreuger, S., & Norman, G. R. (2017). CASPer, an online pre-interview screen for personal/professional characteristics: prediction of national licensure scores. *Advances in Health Sciences Education*, *22*, 327-336.

Galitsky, B., Ilvovsky, D., & Kuznetsov, S. O. (2018). Detecting logical argumentation in text via communicative discourse tree. *Journal of Experimental & Theoretical Artificial Intelligence*, *30*(5), 637-663.

Patterson, F., Zibarras, L., & Ashworth, V. (2016). Situational judgement tests in medical education and training: Research, theory and practice: AMEE Guide No. 100. *Medical teacher*, *38*(1), 3-17.

Schmitt, N., Prasad, J. J., Ryan, A. M., Bradburn, J. C., & Nye, C. D. (2019). Culture as a determinant of option choice in a situational judgement test: A new look. *Journal of Occupational and Organizational Psychology*, *92*(2), 330-351.

Yamaguchi, S. Y., Kanai, S., Kumagai, A., Chijiwa, D., & Kashima, H. (2022). Transfer Learning with Pre-trained Conditional Generative Models. *arXiv preprint arXiv:2204.12833*.

# Effective Evaluations of Teacher-Student Discourse in Classrooms

**Jinnie Shin**
University of Florida
Jinnie.shin@coe.ufl.edu

**Wallace N. Pinto Jr.**
University of Florida
wallace.pintojun@ufl.edu

**Bowen Wang**
University of Florida
Bowen.wang@ufl.edu

**ABSTRACT**: This study explores the evolving role of learning analytics in the context of the shift from product-oriented to process-oriented assessment by evaluating the learning interactions between the students and teacher in the classroom. We examine how transformer models can provide deeper insights into the learning process by evaluating the discourse between the students and the teacher. We illustrate the enhanced classification accuracy achieved through the application of bi-encoders with SBERT (Reimers & Gurevych, 2019) models in assessing teacher-student exchanges that facilitate active meaning-making in mathematics learning.

**Keywords**: classroom discourse, teacher-student talk, Transformers, BERT, SBERT, bi-encoder,

## 1 INTRODUCTION

In the evolving landscape of education, performance feedback emerges as a cornerstone for enriching professional learning among teachers and other educational professionals. Effective feedback, targeting the crucial elements of teaching, must be regular, tailored to individual teachers' needs, and foster reflective practice (Wexler et al., 2020). However, despite significant strides in professional development, augmented by innovative technologies in teacher education, the delivery of timely, personalized, and comprehensive feedback on classroom practices to evaluate the learning practices remains a challenge. A key obstacle in evaluating the effectiveness of various teacher-student interactions is the reliance on predominantly qualitative methods. This study, hence, delves into how learning analytics can bridge this gap. By leveraging data-driven approaches, we explore pathways for providing actionable, context-aware feedback in classrooms, thus enhancing the quality of teacher-student interactions and learning outcomes.

## 2 LITERATURE REVIEW

We build upon the existing literature that explores the training of Transformer-based classifiers for predicting classroom discourse moves (Demszky & Hill, 2023). Our specific focus is on the types of feedback and evaluation pertinent to interactions, which necessitates a detailed examination of both student and teacher discourse. Key elements, such as teacher uptake (defined as a teacher's response to a student's contribution through questions or elaborations; Collins, 1982) and focusing questions (attention to students' thought processes, encouraging clear communication, and fostering reflection on their own and their peers' thoughts; NCTM, 2014), are investigated. These elements have been identified as crucial for enhancing mathematics learning (Alic et al., 2022). Transformer models such as BERT (Devlin et al., 2018) have significantly advanced the analysis of such key discourse elements within classroom interactions. Beyond these large, pre-trained language models, the choice of encoding strategy is crucial. While cross-encoders, which process sentence pairs simultaneously, are widely used, bi-encoders like SBERT present unique advantages. SBERT (Reimers & Gurevych, 2019) independently generates embeddings for each sentence in a pair and then compares these embeddings for semantic similarity, diverging from the traditional approach of transformer models such as BERT, which process single sentences or sentence pairs separately with a focus on the context within individual sentences. This method makes SBERT particularly effective in tasks that require an understanding of semantic similarity. Consequently, leveraging SBERT is potentially more adept at capturing the subtle nuances of student-teacher dialogue, where discerning the semantic relationship between sentences is key (see Fig 1).

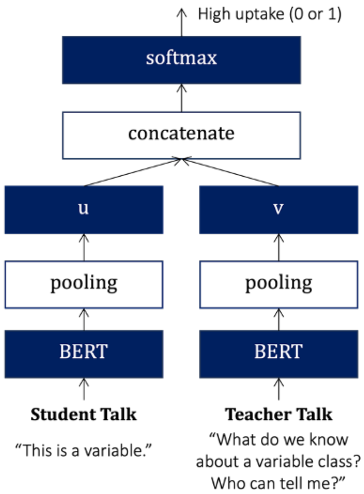



Fig. 1: SBERT architecture

## 3 METHODS

Demszky & Hill (2023) introduced an annotated dataset featuring math classroom interactions, which includes 1,660 transcripts from elementary classrooms. These transcripts were collected by 317 teachers across four schools and were compiled by the National Center for Teacher Effectiveness between 2010 and 2013. Within this dataset, five types of discourse moves were evaluated. Notably, high uptake and focusing questions demonstrated unsatisfactory classification accuracy, as reflected in their F1 scores, when analyzed using their existing approach. We divided into training and validation sets using 5-fold CV, and tokenized using the RoBERTa-base tokenizer for the RoBERTa model and BERT-base-uncased tokenizers for both BERT and SBERT models.

## 4 RESULTS

Table 1 compares the effectiveness of various models, including SBERT, BERT, and RoBERTa, in classifying two types of dialogic discourse moves — high uptake and focusing questions — using precision, recall, and F1 score. Our results obtained with RoBERTa were very close to the baseline model by Demszky & Hill (2023). The SBERT model, with its bi-encoder approach and a batch size of 8, demonstrates superior performance. In the high uptake category, it achieves an F1 score up to 0.78, with the concatenation of $(u, v, |u - v|, u * v)$ and mean pooling. In the focusing question category, SBERT's performance is even more pronounced, with an F1 score reaching up to 0.86 using the product method. In comparison, traditional BERT and RoBERTa models, set at a batch size of 8, show lower effectiveness in the focusing question category, with the highest F1 score for RoBERTa being 0.48. These findings highlight SBERT's enhanced capability in accurately classifying dialogic discourse moves in educational settings, surpassing the performance of traditional cross-encoder models like BERT and RoBERTa.

**Table 1: Classification Performance Results of SBERT, BERT, and RoBERTa Results with a Baseline**

| Model | Concatenate | Pooling | High uptake | | | Focusing question | | |
|---|---|---|---|---|---|---|---|---|
| | | | *Precision* | *Recall* | *F1* | *Precision* | *Recall* | *F1* |
| SBERT [batch size=8] | $u, v, \|u-v\|$ | Mean | 0.73 | 0.74 | 0.73 | 0.86 | 0.83 | 0.85 |
| SBERT [batch size=8] | $u, v, u*v$ | Mean | 0.73 | 0.74 | 0.74 | **0.87** | **0.85** | **0.86** |
| SBERT [batch size=8] | $u, v, u*v, \|u-v\|$ | Mean | **0.77** | **0.78** | **0.78** | 0.72 | 0.85 | 0.78 |
| BERT [batch size=8] | - | - | 0.71 | 0.64 | 0.67 | 0.37 | 0.53 | 0.44 |
| RoBERTa [batch size=8] | - | - | 0.61 | 0.77 | 0.68 | 0.37 | 0.69 | 0.48 |
| RoBERTa [bath size=32] | - | - | 0.69 | 0.67 | 0.68 | 0.48 | 0.45 | 0.46 |
| Demszky & Hill (2023) RoBERTa [batch size=8] | - | - | 0.72 | 0.67 | 0.69 | 0.47 | 0.54 | 0.50 |

## 5 SIGNIFICANCE OF THE STUDY

This study marks a significant advancement in the analysis of classroom interactions through innovative encoding strategies, focusing on Transformer-based classifiers and the bi-encoder model SBERT. Addressing a gap in educational technology, it moves beyond traditional qualitative methods, offering a more scalable and objective data-driven approach. When compared to models like BERT and RoBERTa, the study highlights SBERT's effectiveness in classifying key dialogic discourse moves, such as high uptake and focusing questions. The results showcase SBERT's superior performance and its potential in providing targeted feedback to educators, thus enhancing mathematics instruction and student outcomes. This methodology, scalable across various educational settings, contributes to learning analytics and underscores the transformative role of machine learning in education.

## REFERENCES


Alic, S., Demszky, D., Mancenido, Z., Liu, J., Hill, H., & Jurafsky, D. (2022). Computationally identifying funneling and focusing questions in classroom discourse. *arXiv preprint arXiv:2208.04715*.

Collins, J. (1982). Discourse style, classroom interaction and differential treatment. *Journal of Reading Behavior, 14*(4):429–437.

Demszky, D., Liu, J., Mancenido, Z., Cohen, J., Hill, H., Jurafsky, D., & Hashimoto, T. (2021). Measuring conversational uptake: A case study on student-teacher interactions. *arXiv preprint arXiv:2106.03873*.

Demszky, D., & Hill, H. (2023). The NCTE Transcripts: A dataset of elementary math classroom transcripts. *arXiv preprint arXiv:2211.11772*.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Humeau, S., Shuster, K., Lachaux, M. A., & Weston, J. (2019). Poly-encoders: Transformer architectures and pre-training strategies for fast and accurate multi-sentence scoring. *arXiv preprint arXiv:1905.01969*.

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.

Wexler, L. J. (2020). How feedback from mentor teachers sustained student teachers through their first year of teaching. *Action in Teacher Education, 42*(2), 167-185.

# Exploring the Relationship Between Pre-service Teachers' Uses of Generative AI, Questioning Skills, and Teacher Knowledge

**Pei-Yi Lin**
Department of Education and Learning Technology, National Tsing Hua University, Taiwan
peiyilin@mx.nthu.edu.tw

**ABSTRACT**: Questioning is a crucial method to assess students' understanding in class. In the AI age, how teachers use AI tools to design learning activities to promote students' thinking needs more research and sufficient evidence. This study investigated how pre-service teachers use ChatGPT to generate high-level cognitive questions for students. Moreover, this study also investigated their perceptions when they co-regulated learning with AI and their technology-related teacher knowledge. The participants were 14 pre-service teachers enrolled in an interdisciplinary learning design course. In this course, they engaged in design activities to design questions to provoke students' thinking and deepen understanding: (1) They decided on an interdisciplinary learning theme to develop learning activities and questions to promote students' learning. (2) They used ChatGPT to produce questions according to their designed learning activities. (3) The participants referred to the AI-generated questions to revise and refine their questions. The results showed that the participants provided a significantly greater number of high-cognitive questions after using ChatGPT, and their co-regulated learning with AI was significantly related to technological content knowledge and technological pedagogical content knowledge. Understanding how teachers use AI to provide high-level questions is critical for improving teacher support and preparation.

**Keywords**: generative artificial intelligence, questioning, co-regulated learning, teacher knowledge

## 1 RESEARCH CONTEXT

Asking questions is one of the most critical pedagogical techniques to inspire students to think and inquire. Thus, developing and promoting teachers' capacity to formulate meaningful questions to uncover students' learning is crucial. The advent of artificial intelligence (AI) agents allows teachers to practice asking AI questions and ask AI to refine their questions in order to optimize their questioning skills. For example, Lee and Yeo (2022) used conversational chatbots to simulate students' responses and interactions to train pre-service teachers' responsive teaching skills. Generative AI (GAI), e.g., ChatGPT, can assist teachers in developing professional knowledge and skills, such as preparing lesson plans and activities (Kasneci et al., 2023) and designing pedagogy (Bhat et al., 2022).

When people use ChatGPT, they must identify and use the information sources well. Effective SRL is essential for users to make conscious choices while working toward learning objectives in an AI-supported environment (Winne & Perry, 2000). Empirical studies found that AI technologies can provide instant support for learners' SRL processes (Kay, 2023). Nonetheless, few studies examine learners' role in regulating their uses of AI chatbots and their relation to learners' cognitive performance. Examining how learners use meta-cognitive processes such as self-explaining, monitoring, reflection, and planning the retrieved data from ChatGPT is necessary.

This study aimed to facilitate pre-service teachers' questioning ability with an intervention using ChatGPT. The pre-service teachers design 6E (engage, explore, explain. engineer, enrich, and evaluate) learning activities and questions to provoke pupils' thinking and knowledge construction in their lesson plans. The first research question of this study was: Did the cognitive level of participants' question designs improve after using ChatGPT? The second research question was: Did the participants' co-regulated learning with ChatGPT relate to their technology-related teacher knowledge?

# 2 METHOD

## 2.1 Context and Participants

The study took place in a graduate-level interdisciplinary learning design course at a university in Taiwan with 14 pre-service teachers (50% male). The 6E learning design activities crossed seven weeks (the sixth week was a national holiday, so no class) and three periods per week. Each week, lectures and group discussions were based on the weekly assigned papers for two periods; the participants engaged in design activities and presented their discussion results for one period. In the fourth week, each group reported their group's design of 6E learning questions. The intervention was implemented in the fifth week. The pre-service teachers reported that when they used ChatGPT, they first asked GPT-3.5 to explain 6E learning, then input their 6E learning goals and activities to tell GPT to generate 6E learning questions that met their 6E learning goals and activities. Then, they designed and revised based on GPT-generated 6E learning questions for the final presentation in the seventh week.

## 2.2 Data Sources

Data includes qualitative and quantitative parts. Qualitative data were retrieved from pre-service teachers' pre-and post-6E learning question designs in the fourth and seventh weeks. The author analyzed the pre- and post-interdisciplinary 6E learning questions based on the revised Bloom's six levels of cognitive taxonomy (Anderson et al., 2001). The six cognitive domains are further categorized into lower-order (i.e., remember, understand, apply) or higher-order (i.e., analyze, evaluate, create) thinking. To ensure accuracy in the categorization, a total of ninety question prompts were independently classified by two researchers and then compared for any differences. The inter-rater reliability was 0.90, showing that the coding was reliable. Quantitative data was retrieved from the participants' self-reported surveys about their perceptions of co-regulated learning with ChatGPT to support their questioning designs and technology-related knowledge (i.e., technological content knowledge (TCK), technological pedagogical knowledge (TPK), and technological pedagogical content knowledge (TPACK)) questionnaire. The co-regulated learning and TPACK surveys were adapted from validated studies by Sha et al. (2012) and Chai et al. (2011).

# 3 RESULTS

The first research question examines whether teacher questioning practices improved after the pre-service teachers' use of ChatGPT. The author analyzed their designed questions and conducted a paired sample t-test to compare their pre- and post-question design changes—the results of t-tests were conducted on the coding for each question. The results showed that their low-cognitive questions decreased (pre-test: $M$ = 5.75, $SD$ = 1.89, post-test: $M$ = 3.75, $SD$ = 0.96, $t$ = 1.48, $p$ = 0.24), while their high-cognitive questions significantly increased (pre-test: $M$ = 3.75, $SD$ = 2.50, post-test: $M$ = 9.25, $SD$ = 2.87, $t$ = -5.28, $p$ < 0.05). There were notable changes in that the pre-service teachers tended to ask more high-cognitive questions in their final 6E learning micro-teaching plans after referencing AI-generated questions. In addition, the author also counted the number of new questions that were added to the post-6E learning questions. There were 22 low-level and 52 high-level cognitive questions. These numbers are evidence of substantial improvement in teacher questioning practices in this study. For example, one of the group's pre-6E learning questions in the explaining phase was, "Why is the tail important for some animals to survive or live?" This was categorized to the "understand" level. They revised the questions: "Do humans have no tails? What would happen if humans kept their healthy tails?" which was labeled to the "evaluate" level after referencing ChatGPT.

To answer the second research question, this study surveyed the participants' perception of co-regulated learning with AI and technology-related knowledge, i.e., TCK, TPK, and TPACK. The coefficients of the Spearman's rho correlation analysis results indicated that their co-regulated learning with ChatGPT was significantly correlated to TCK ($\rho$ = 0.76, $p$ < .01) and TPACK ($\rho$ = 0.55, $p$ < .05). However, there was no significant correlation to TPK ($\rho$ = 0.27, $p$ = .35).

This study was implemented as a preliminary study to show that the uses of ChatGPT could advance the pre-service teachers' questioning design and their co-regulated learning with GAI related to their TCK and TPACK. There were also some limitations. This study only had 14 participants, and the design activities were conducted as part of the course tasks. Future studies should have more participants and use GAI to assist pre-service teachers in designing different teaching subject matter activities.

## REFERENCES

Anderson, L. W., Krathwohl, D. R., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., … & Wittrock, M. (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy*. New York: Longman Publishing.

Bhat, S., Nguyen, H. A., Moore, S., Stamper, J., Sakr, M., & Nyberg, E. (2022). Towards automated generation and evaluation of questions in educational domains. In *Proceedings of the 15th International Conference on Educational Data Mining* (Vol. 701).

Chai, C. S., Koh, J. H. L., & Tsai, C. C. (2011). Exploring the factor structure of the constructs of technological, pedagogical, content knowledge (TPACK). *Asia-Pacific Education Researcher, 20*(3), 595-603.

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and individual differences, 103, 102274.

Kay, J. (2023). Foundations for Human-AI teaming for self-regulated learning with explainable AI (XAI). *Computers in Human Behavior*, 107848.

Lee, D., & Yeo, S. (2022). Developing an AI-based chatbot for practicing responsive teaching in mathematics. *Computers & Education, p. 191*, 104646.

Sha, L., Looi, C. K., Chen, W., & Zhang, B. H. (2012). Understanding mobile learning from the perspective of self-regulated learning. *Journal of computer assisted learning*, *28*(4), 366–378.

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In *Handbook of self-regulation* (pp. 531–566). Academic Press.

# Analyzing L2 Learner-ChatGPT Interaction in English Essay Writing with Epistemic Network Analysis

**Hyeji Jang, Lingxi Jin, Hyo-Jeong So**
Ewha Womans University
hyejijang@ewhain.net, jinlingxi@ewhain.net, hyojeongso@ewha.ac.kr

**ABSTRACT**: This study aims to examine the L2 learners' interactions with ChatGPT in the essay writing process, particularly identifying distinctive patterns across varying proficiency groups. Ten Korean undergraduate students were divided into intermediate and advanced learner groups and completed a TOEFL essay task using ChatGPT 4. The conversation logs were analyzed using the learner-ChatGPT Interaction coding framework in the four stages of writing: prewriting, drafting, revising, and editing. Epistemic Network Analysis (ENA) was employed to visualize interaction patterns in the two groups. Overall, the results indicate that the advanced group frequently sought information regarding the essay topic during the prewriting and drafting phases. In contrast, the intermediate group showed a distinct focus, primarily soliciting assistance for suitable words, phrases, or sentence expressions during the drafting phase. This study highlights the necessity for customizable approaches that address the requirements of learners at different proficiency levels, contributing to the ongoing research on the effective integration of Generative AI in language education.

**Keywords**: Generative artificial intelligence, ChatGPT, L2 learners, Epistemic Network Analysis

## 1 INTRODUCTION

Generative AI (GenAI), exemplified by platforms like ChatGPT, has demonstrated its potential in assisting L2 (second-language) learners in writing tasks, leveraging automatic text generation and translation capabilities. However, the current applications of ChatGPT in L2 writing lack sufficient pedagogical considerations, specifically in terms of providing writing stage-specific scaffolding and addressing learners' diverse competency levels. Existing literature underscores the variation in learners' writing focuses and needs across proficiency levels; for instance, intermediate learners prioritize grammar while advanced learners focus on meaning-making (Such, 2021). Consequently, understanding how L2 learners of different proficiency levels interact with GenAI during the writing process is essential to inform the development of personalized pedagogical support. To address this gap, this work-in-progress research aims to investigate ChatGPT-learner interactions in the essay writing process, discerning patterns between different proficiency groups.

## 2 METHOD

### 2.1 Research Context and Participants

Ten Korean undergraduate students (7 females and 3 males) were recruited through a convenient sampling method for participation in this study. The students were categorized into two proficiency groups, namely intermediate learners (CEFR Level B) and advanced learners (CEFR Level C), determined by their Common European Framework of Reference for Languages (CEFR) levels. Five learners were assigned to each group. They were asked to complete a TOEFL writing task within a 30-minute timeframe. The task required the generation of a 300-word essay expressing agreement or disagreement with the statement "Television advertising directed toward young children (aged two to five) should not be allowed." To provide writing stage-specific assistance, the 'Custom Instructions' feature in ChatGPT 4 was employed, which involved the insertion of tailored prompts to guide learners through the four stages of the writing process, namely pre-writing, drafting, revising, and editing.

### 2.2 Data Analysis

In the lab setting, the participants' use of ChatGPT was screen-recorded, and conversation logs were collected for interaction analysis. Based on Laksmi's (2006) writing process approach, a coding framework (Table 1) was developed with some modifications. Two independent coders analyzed the data, achieving a satisfactory level of agreement (Cohen’s kappa coefficient of .74). Any disparities between the coders were reconciled through iterative discussions until a consensus was reached. For the subsequent data analysis, Epistemic Network Analysis (ENA) by Shaffer et al. (2016) was employed to visualize interaction patterns and to facilitate comparisons between the networks of the two groups. The ENA Web Tool was utilized, configuring the model with units representing both groups, each comprising five learners. The learners' text logs within ChatGPT were designated as the conversation, with a window size of 5 to align with the five learners in each group. Given the relatively small sample size, we conducted the Mann-Whitney U test to examine statistical differences between the groups along the axis projecting networks into a two-dimensional space.

**Table 1: Coding framework for learner-ChatGPT interaction in writing.**

| Writing stage | Code description | | | |
|---|---|---|---|---|
| Stage 1: Prewriting | PI | Prewriting-organizing information | PIC | Prewriting-gathering information |
| | PS | Prewriting-defining a topic sentence | PW | Prewriting-asking about words in English |
| | PO | Prewriting-writing an outline | PCT | Prewriting-asking to translate |
| Stage 2: Drafting | DC | Drafting-emphasizing content | DP | Drafting- asking about phrases in English |
| | DIC | Drafting- gathering information | DS | Drafting- asking about sentences in English |
| | DW | Drafting- asking about words in English | | |
| Stage 3: Revising | RSC | Revising-sharing writings | RPC | Revising-participating in discussions |
| | RCC | Revising-changing the compositions | | |
| Stage 4: Editing | EPC | Editing-proofreading the writing | ECC | Editing-correcting mechanical errors |

## 3 RESULTS

Each group’s interaction networks are presented in Figure 1. The model showed that MR1 (means rotation) and SVD2 (singular value decomposition) dimensions accounted for 13.0% and 19.5% of the variance in the data, respectively. The Mann-Whitney U test ($U$=25.00, $p$<.01, $r$=-1.00) revealed significant differences between the advanced learner (Mdn=-0.38, $N$=5) and the intermediate learner (Mdn=0.32, $N$=5) groups along the X-axis (MR1). On the other hand, no significant differences were observed in the Y-axis (SVD2). In the ENA (Figure 1), the advanced learner group showed strong connections by centering information gathering in prewriting (PIC) and drafting (DIC). The strong connections with PIC and DIC are asking for appropriate words in English in drafting (DW), outlining in prewriting (PO), and correcting mechanical errors (ECC). Meanwhile, the intermediate learner group displayed robust associations with asking for appropriate words in English in drafting (DW). The strong connections with DW are asking for appropriate phrases in English in drafting (DP), gathering information in prewriting (PIC), correcting mechanical errors with ChatGPT (ECC), and asking for appropriate sentences in drafting (DS).

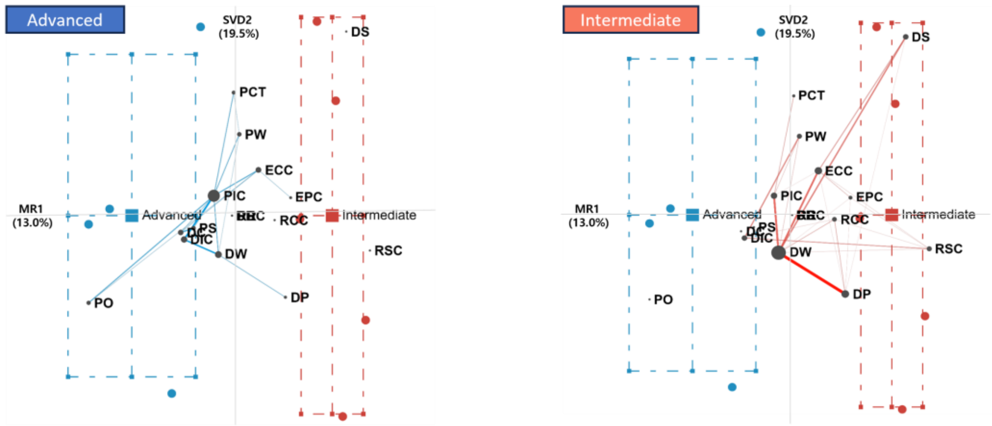



**Figure 1: ENA projection graphs of the advanced and intermediate learner groups**

## 4 DISCUSSION, LIMITATIONS, AND FUTURE DIRECTIONS

In this study, two predominant patterns of ChatGPT-L2 learner interaction were identified: (a) asking for information pertaining to the essay topic during prewriting and drafting, and (b) asking assistance for appropriate words, phrases, and sentence expressions during drafting, along with addressing mechanical errors in the editing phase. Moreover, significant differences in the ENA were identified between the intermediate and advanced groups. Specifically, in the advanced group, learners frequently sought information on the essay topic during prewriting and drafting, indicating a proactive approach to idea development and outline creation. Conversely, in the intermediate group, the primary emphasis was on requesting suitable words, phrases, or sentence expressions during the drafting phase. These findings underscore the importance of customizing GenAI tools for L2 writing to accommodate the distinct needs observed in these two groups. For example, a GenAI tool designed for L2 writing could offer prompts for idea generations and provide learners with cognitive diagnostic scaffolding (Yang et al., 2023). We acknowledge the limitations of the present study, including the small sample size and the absence of findings related to beginning-level learners. Recognizing these limitations, subsequent iterations of this work-in-progress aim to present more comprehensive findings by expanding the dataset across various proficiency levels of L2 learners.

## REFERENCES

Laksmi, E. D. (2006). Scaffolding students' writing in EFL class: Implementing process approach. *Teflin Journal, 17*(2), 144-156. https://doi.org/10.15639/teflinjournal.v17i2/152-165

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics, 3*(3), 9-45. https://doi.org/10.18608/jla.2016.33.3

Such, B. (2021). Scaffolding English language learners for online collaborative writing activities. *Interactive Learning Environments, 29*(3), 473-481. https://doi.org/10.1080/10494820.2019.1579233

Yang, G., Zhou, W., Zhou, H., Li, J., Chen, X., & Tu, Y. F. (2023). An empirical study of the effects of intelligent cognitive diagnostic feedback strategy on L2 writing performance, epistemic structure, and transferability. *Education and Information Technologies*, 1-34. https://doi.org/10.1007/s10639-023-11905-3

# Exploring Learning Analytics Approach Using Log Data of Programming Environment Integrating Generative AI

**Manabu Higashida[1*], David Soto[1], Shizuka Shirai[1], Mayumi Ueda[2]**
[1]Osaka University, [2]University of Marketing and Distribution Sciences
*manabu.higashida.cmc@osaka-u.ac.jp

**ABSTRACT**: This poster describes the pilot study exploring the learning analytics (LA) approach for programming education using Generative AI. We developed a proprietary Learning Record Store (LRS) that can record responses between students, and the interactive programming language and between students and OpenAI mediated and collected data in a first-year university programming class. The results of the qualitative analysis using collecting data showed that the proposed approach using student interaction patterns is a promising way to understand students' learning process.

**Keywords**: Programming education, lrs, knowledge tracing, large language model

## 1 INTRODUCTION

The emergence of ChatGPT and its application to the field of programming assistance has been gaining popularity and attention from its early stages. However, most research focuses on the use and value (e.g., Biswas, 2023; Chen et al., 2023); there is little research on learning analytics with learning logs using Generative AI. Understanding students' learning processes and outcomes using generative AI requires a new framework to determine how students have modified program codes through interactions with the generative AI. This study aimed to explore a learning analytics (LA) approach, considering the interaction with Generative AI. We report the results that evaluate the progress of program code by extracting code from cell program codes and generative AI prompt responses while pointing out the differences between the codes.

## 2 METHODOLOGY

Firstly, we have incorporated the programming assistance capabilities of generative AI (ChatGPT-3.5) into JupyterHub and developed a proprietary Learning Record Store (LRS) specializing in programming education support. Our LRS can record responses for all student programming activities, including cell inputs, outputs, and error notifications. We have incorporated the functionality to record responses from generative AI (prompts and responses) in conjunction with this. Then, we collected logs produced while students developed program code by interacting with generative AI in the programming course for first-year university students.

In our course, we explore the evolution of numerical notations. In the 2023 academic year, students were assigned the task of developing program codes that convert decimal numbers into Roman numerals and other formats through responses with generative AI. Concurrently, we conducted exercises for students to respond to fill-in-the-blank programming problems, also using generative AI responses. Subsequently, as part of the final achievement assessment, we tasked students with solving a problem that involved converting decimal numbers to Chinese numerals. To explore an LA approach, considering the interaction with Generative AI, we analyze the relationship between the final achievement assessment and the following five evaluation criteria. These evaluation criteria were considered with the aim of assessing the comprehensive skills of students while interacting with generative AI within the context of programming prompts and responses:

**E1. Ability to Devise Appropriate Prompts:** capacity to create suitable prompts for generative AI.
**E2. Generation of Prompts Encouraging Code Refinement in Responses:** Proficiency in generating prompts that encourage the modification of program code included in responses.
**E3. Generation of AI-Friendly Code Skeletons:** The skill to generate AI-friendly code skeletons by providing program code included in fill-in-the-blank problems to the generative AI.
**E4. Assessment of Current Quality of Generative AI Responses:** The capability to discern the quality of the generative AI responses and formulate alternative suitable prompts, and the generation of prompts based on the program code included in fill-in-the-blank problems.
**E5. Execution and Verification of Code in Responses:** The ability to execute and verify program code included in the responses.

## 3 RESULTS

First, regarding the E1, only 2 out of 14 students demonstrated the ability to devise appropriate prompts by incorporating detailed specifications discussed in the pre-guided problem explanations. Another four students issued responses based on the given problem statement as the specification. Figure 1 is density plots of these four students, with the horizontal axis being the time-lapse, 50 minutes. Small spots at the top are prompts to the AI, and the lower part is the input/output and response of the Jupyter Notebook cell operation. Note that large yellow circles indicate the occurrence of error. The remaining eight students did not receive specifications, where four attempted to write the program code from scratch, and the other 4 provided the fill-in-the-blank program code as is without generating their own prompts (Figure 4).

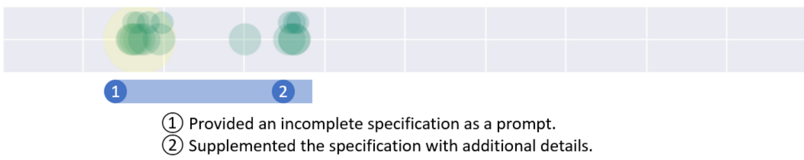



**Figure 1: Case of providing an incomplete specification as a prompt**

Next, the results of E2, although 3 out of 14 students were able to prompt a response modification in the response of the generative AI, 2 of them had reached the input token limit without being able to identify the limits of the Generative AI at the moment (Figure 2). The remaining students had outstanding abilities that are noted later.

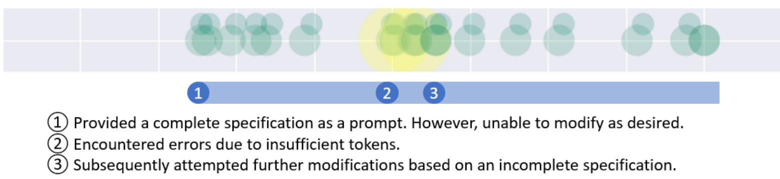



**Figure 2: Case of providing a complete specification as a prompt**

Then, concerning E3, six of the 14 students had been providing generative AI with fill-in-the-blank problems as direct prompts, thus receiving responses that may include errors (Figure 3). However, none of these students demonstrated the ability to conduct further verification and troubleshooting. Due to a predominant focus on engaging with generative AI responses, these students did not develop skills in fundamental trial-and-error programming and were unable to acquire proficiency in verifying potential inaccuracies present in generative AI responses.

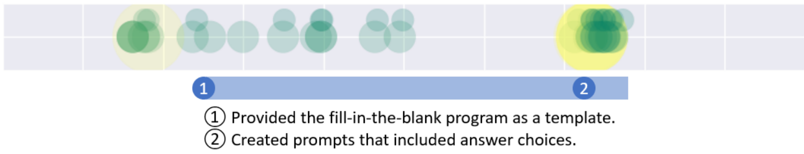



**Figure 3: Case of providing the fill-in-the-blank as a template in prompt**

Finally, regarding E4 and 5, one student was deemed capable of providing the most outstanding responses. This student not only supplied detailed specifications to the generative AI, but also assessed whether the generated responses could be applied as solutions to the given problems. Additionally, as an alternative approach, the student provided a fill-in-the-blank code skeleton and prompted the generative AI to correct any inaccuracies in its responses. It is noteworthy, however, that this student did not execute the final program code, and their programming skills may be deduced from their ability to identify errors in generative AI responses based on past experiences.

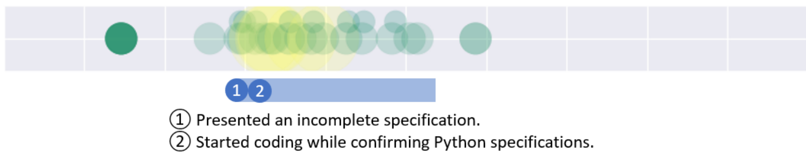



**Figure 4: Case of own coding while confirming Python specification in Prompts**

## 4 CONCLUSION AND FUTURE WORK

The results of the qualitative analysis using collecting data showed that the proposed approach using student interaction patterns is a promising way to understand students' learning process. Based on the results of this experiment and considering the outstanding response as a model, efforts are underway to revise teaching materials and enhance the LRS in the subsequent academic years. Some of the proposed improvements include: (1) Omitting the fundamentals of programming so that students can focus on logical thinking and use generative AI efficiently; (2) Ensuring that problem statements maintain an ample context to avoid reaching the token limit for the response generation; (3) Validating the acquired final program code and prompting generative AI for verification of the results as well; and (4) raising awareness of the ability to formulate prompts based on academic knowledge as part of remedial education.

## REFERENCES

Biswas, S. (2023). Role of ChatGPT in Computer Programming.: ChatGPT in Computer Programming. *Mesopotamian Journal of Computer Science*, *2023*, 8-16.

Chen, E., Huang, R., Chen, H., Tseng, Y., & Li, L. (2023). GPTutor: a ChatGPT-powered programming tool for code explanation. *International Conference on Artificial Intelligence in Education*.

# Understanding teachers' perspectives on ethical concerns and skills to use AI tools

**Ayaz Karimov**
University of Jyväskylä
akarimov@jyu.fi

**Mirka Saarela**
University of Jyväskylä
mirka.saarela@jyu.fi

**Tommi Kärkkäinen**
University of Jyväskylä
tommi.karkkainen@jyu.fi

**ABSTRACT**: This poster paper explores teachers' perspectives on the integration of artificial intelligence (AI) tools in education, focusing on ethical concerns and the requisite skills for effective implementation. A survey of 48 educators from 32 countries revealed widespread familiarity with AI (92%) and personal use (85%). Ethical concerns, notably about decision control and biases, were expressed by 57.5% of participants. The majority (90%) expressed interest in AI training, highlighting a need for professional development. Educators emphasized the importance of possessing practical skills in lesson preparation, student engagement, and prompt engineering. Moreover, it is important to note that the insights presented in this study are derived from self-reported knowledge and experiences of educators, providing a subjective perspective on their engagement with AI tools in education.

**Keywords**: AI ethics, AI tools, AI skills, AI in education, AI literacy

## INTRODUCTION

While there are innovative emerging technologies and tools in the education field, during the last five years, most educators have encountered challenges in keeping up with these updates. Because each new technology requires time to learn and find the most efficient way to implement it in the classroom, this applies particularly to the great advancements in artificial intelligence (AI) (Karimov et al., 2023). AI applications in education have shown promise in personalizing learning experiences, enhancing student engagement, and optimizing learning outcomes (Magomadov, 2020). Within the implementation of AI tools, the educators' role is important since they are one of the main stakeholders in the learning process (Celik et al., 2022). Prior research has primarily concentrated on understanding teachers' perspectives on specific AI tools, often overlooking their broader motivations, skills, and concerns associated with these AI tools (AlAfnan et al., 2023). Nonetheless, the integration of AI tools into educational settings is a complex field influenced by the different reasons and concerns of educators (Chounta et al., 2022). In this study, our main focus was to identify the key ethical concerns of educators and explore the skills they believe are necessary for effectively incorporating AI tools into their teaching methods.

## METHODOLOGY

To understand teachers' perspectives on the integration of AI tools in their teaching process, we developed a survey that was shared across 54 teacher communities on social media platforms such as Facebook and LinkedIn, garnering responses from 48 educators spanning 32 different countries (Figure 1). The respondents had varying levels of teaching experience, with the majority (78.3%) having more than 10 years of teaching experience. Almost all respondents (92%) indicated that they are familiar with the concept of AI in education (Table 1). Moreover, 85% of educators reported that they have personally used AI tools or applications as part of their teaching. Furthermore, the participants' diverse teaching backgrounds cover a broad array of subjects, including science, language, technology education, social studies, computer studies, mathematics, and literature.

**Figure 1: Participants' country demographics**

**Table 1: Summary of survey responses regarding ethical concerns, interest in training, familiarity and utilization of AI tools in education**

| Question | Yes (%) | No (%) | Maybe (%) |
|---|---|---|---|
| Familiarity with AI in education | 92 | 8 | 0 |
| Use of AI tools in education | 60 | 40 | 0 |
| Concerns about ethical implications | 57.5 | 22.5 | 20 |
| Interest in receiving AI training | 90 | 0 | 10 |

## RESULTS

In the initial section of the survey, teachers detailed the AI tools they employ and the driving factors behind their utilization. Teachers mentioned that they utilize different AI tools such as ChatGPT, AIforslides, Canva, Huggingface, Curipod, and Grammarly. 90% of participants expressed interest in receiving training or professional development related to the use of AI tools in education, while 10% are uncertain or may consider it, and none indicated a lack of interest. Moreover, the responses to our research question regarding the topics teachers are interested in for AI as required skill reveal a diverse range of preferences. Educators expressed interest in practical applications of AI, such as utilizing it to teach languages and integrating it into the science curriculum for enhanced learning and engagement. Notably, participants seek to learn the functionalities of AI tools, emphasizing their application in the humanistic and scientific domains. Beyond the application spectrum, teachers are interested in the broader context of AI in education, with some highlighting the importance of inclusive teaching and the potential role of AI in making education more accessible. Furthermore, educators demonstrate a practical outlook by expressing interest in topics related to lesson preparation, student engagement, and prompt engineering. This suggests a desire for hands-on knowledge and strategies for integrating AI seamlessly into day-to-day teaching practices. Moreover, there is an interest in fostering student-AI interaction within the classroom, underlining educators' recognition of AI as a tool for formative evaluation and engagement.

The answers to the question about ethical worries related to using AI in education show that many participants are aware of and concerned about this issue. Approximately 70% of respondents express clear concerns about the ethical implications of using AI in education, with 20% indicating uncertainty or a "Maybe" response. In contrast, 20% of participants explicitly stated they have no concerns about the ethical aspects of AI integration in education. A major worry is the perceived lack of control over

AI decisions (see, e.g., Saarela et al., 2021), with 19 teachers sharing this concern. Close behind is the issue of bias in AI algorithms, mentioned by 18 educators who fear potential biases impacting educational outcomes. Furthermore, 11 teachers highlighted the importance of understanding how AI systems make decisions, expressing reservations about their interpretability. The exclusion of practitioner expertise is a significant concern for 16 respondents, indicating worries about sidelining the valuable insights and experiences of educators. Other concerns raised by individual participants include worries about the accuracy of AI tools, student access, lack of experience in using AI, privacy concerns involving data collection and age restrictions, potential loss of original thought, societal division due to AI utilization, and overlooking learning aspects like autonomy and socialization.

## CONCLUSION

The study emphasizes educators' strong interest (90%) in AI training, underscoring the need for further professional development. Educators stress the importance of practical skills for effective AI integration into teaching, particularly in lesson preparation, student engagement, and prompt engineering. Ethical concerns related to AI implementation are evident, with 57.5% expressing worries. Key concerns encompass issues of control over AI decisions, biases in algorithms, and potential exclusion of practitioner expertise. Subsequently, we would like to highlight the importance of introducing the concept of AI for educators. Therefore, as a preliminary step before delving into the discussion of core AI skills, it is crucial to provide educators with foundational AI literacy skills.

## ACKNOWLEDGEMENT

The collaborative efforts of the Swiss Cyber Institute are gratefully acknowledged, as their support significantly enhanced the quality and depth of this research. The work was also supported by the Academy of Finland (grant no. 356314).

## REFERENCES

AlAfnan, M. A., Dishari, S., Jovic, M., & Lomidze, K. (2023). Chatgpt as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses. Journal of Artificial Intelligence and Technology, 3(2), 60-68.

Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. TechTrends, 66(4), 616-630.

Chounta, IA., Bardone, E., Raudsep, A. Exploring Teachers' Perceptions of Artificial Intelligence as a Tool to Support their Practice in Estonian K-12 Education. Int J Artif Intell Educ 32, 725–755 (2022).

Karimov, A., Saarela, M., & Kärkkäinen, T. (2023). The impact of online educational platform on students' motivation and grades: the case of Khan Academy in the under-resourced communities.

Magomadov, V. S. (2020, November). The application of artificial intelligence and Big Data analytics in personalized learning. In Journal of Physics: Conference Series (Vol. 1691, No. 1, p. 012169). IOP Publishing.

Saarela, M., Heilala, V., Jääskelä, P., Rantakaulio, A., & Kärkkäinen, T. (2021). Explainable student agency analytics. IEEE Access, 9, 137444-137459.

# Unpacking the Elicitation Process for Multimodal Learning Analytics based Feedback

**Daniel Sánchez, Rogers Kaliisa, Rachelle Esterhazy, Crina Damsa**
University of Oslo, Department of Education
{daniel.sanchez, rogers.kaliisa, rachelle.esterhazy, crina.damsa}@iped.uio.no

**ABSTRACT**: This study introduces a method to enhance Multimodal Learning Analytics (MMLA) feedback by integrating expert analysis with multimodal data to generate feedback that aligns with pedagogical principles and professional standards. Conducted in a nursing simulation at a Norwegian university, the study employs Design-Based Research, focusing on identifying key learning incidents through multimodal data. The findings emphasize the need for context-sensitive, expert-driven, and purposeful data collection in MMLA tool development.

**Keywords**: Multimodal Learning Analytics, Feedback, Simulation-based Learning

## 1 INTRODUCTION

The imperative for integrating multiple data streams in educational settings is becoming increasingly evident. Multimodal Learning Analytics (MMLA) uses advanced computational techniques to analyze such data to enhance learning processes (Ochoa, 2022). However, there is an overemphasizing of the technological aspects, neglecting the pedagogical and subject-domain principles (Munshi & Deneen, 2018), and a challenge in comprehensively interpreting and analyzing data from various modalities without oversimplification. Building upon Ochoa (2022), this poster moves beyond traditional MMLA's focus on predicting learning performance. It introduces an elicitation process to refine data selection and interpretation, enhancing feedback effectiveness, particularly in simulation-based learning.

## 2 DESIGN CONTEXT & PROCESS

This design process in the current research aims to develop a tool to provide MMLA-based feedback to graduate nursing students in simulated Intense Care Unit (ICU) at a Norwegian university. The study employs a Design-Based Research (DBR) approach, where nursing teachers, engineers, and researchers actively engage in an iterative process to design the tool. The ICU simulation starts with a briefing, then students enact the scenario and conclude with a debriefing on the scenario's events. During the first DBR cycle, data were collected from co-design meetings engaging five teachers and from simulation sessions with 22 graduate students who provided informed consent. This poster explains the elicitation-related aspects and their application during this first iteration cycle and for illustrative purposes center on a multimodal feedback input developed related to stress response.

## 3 LEARNING ELICITATION IN MMLA

The Learning elicitation in MMLA, a key process for understanding the learning experience, involves understanding data about learners' interactions, behaviors, and the learning context. This process guides which multimodal data should be collected and how it should be utilized and presented, thus informing the further selection of sensors, analytics, and feedback mechanisms. Based on literature, the MMLA tool development is structured around the following set of principles and process steps:

### 3.1 Principles

*Principle 1: Context Dependency of Multimodal Data.* The value and interpretation of multimodal data are intrinsically linked to the context of learner interactions within educational environments (Wise, 2019). This principle is essential for accurately analyzing and meaningfully interpreting the data, ensuring that each learner's action or behavior is understood within its specific context.

*Principle 2: Expert-Driven Feedback.* This principle stresses the importance of subject-matter experts (SMEs) in interpreting context-dependent data. SMEs can identify specific actions and behaviors crucial for professional practice, which novices may miss (Ericsson, 2008). Their expertise supports accurately interpreting student actions and providing relevant, professionally-aligned feedback.

*Principle 3: Purposeful Data Collection.* From the onset, it is crucial to define what MMLA data will be gathered and how it will be presented to students, ensuring alignment with established practices and learning objectives (Wise, 2019). Drawing inspiration from elicited expertise and seeking to emulate it where feasible, informs a deliberate planning process for selecting sensors, analytics, and feedback mechanisms in the development of the MMLA tool.

### 3.2 Elicitation Process

**Step 1: In-Depth Contextual Elicitation.** To fully unpack the context of the learning activity, the simulation scenario's design was closely examined, along with observations of the environment and student engagement. Analysis revealed that the scenario targeted learning outcomes in communication, teamwork, and patient assessment, with pivotal moments impacting its dynamics. Observations of debriefings highlighted a pedagogical approach incorporating guided inquiry, where teacher facilitated students' reflection about their performance. A learning outcome identified was the stress response, influencing communication and teamwork in the ICU. These findings were shared with SMEs to ensure a proper understanding of the learning design context.

**Step 2: Incidents Identification.** Following the expert-driven approach, in this step, SMEs identified key learner actions and behaviors, termed 'incidents,' in the ICU scenario. Through the analysis of video recordings of students in this setting, SMEs highlighted specific observed aspects and relevance per incident. This process found potential ways to identify such incidents based on cues like body motion, verbal interaction, and changes in speech intonation, which were then linked to learning outcomes. Further analysis involved correlating the incidents with feedback obtained from debriefing comments by teachers and the associated incidents during simulation. This way, it was possible to (1) identify a set of incidents, explore their pedagogical relevance and the modalities involved in their detection, and (2) connect these incidents with potential feedback comments.

**Step 3: Integration with Literature.** This final step contrasted previous findings with literature to establish a robust foundation for choosing technologies and analytics suited for incident tracking in the learning environment. For instance, the literature review provided insights into how stress impacts simulations and healthcare practices, guiding the choice of MMLA data. This MMLA data includes physiological indicators, such as heart rate, to detect stress levels during simulations. About the context, the literature validates debriefing as a pedagogical approach common in healthcare. This approach ensures the strategic selection of sensors and analytics that capture essential data and the feedback mechanism to enhance student reflection using the MMLA tool.

## 4 MMLA-BASED FEEDBACK

The elicitation process has been instrumental in shaping MMLA-based feedback. Figure 1 demonstrates the integration of elicited contextual knowledge by displaying pivotal moments of a simulation on a timeline, marked by vertical green-dotted lines and yellow areas. The upper graph captures incidents where the heart rate exceeds the baseline, as indicated by the red dotted line. Complementing this, the lower graph shows body movement (net acceleration), aiding in discerning whether the stress response is due to physical exertion or predominantly psychological. In this setting, the combination of multimodal data and prompts designed to initiate reflective thinking among students is what becomes into MMLA-based feedback. For example, highlighting consistent heart rate peaks across all students with a comment like “Stress is usual in this scenario” stimulates discussions about stress management in the ICU. Similarly, individual stress peaks, when paired with corresponding video and observations such as "one can be stressed inwardly, but appear calm outwardly," offer personalized feedback. This elicitation process contributed to a meaningful, grounded in evidence feedback, aligned with pedagogical practices and learning objectives.

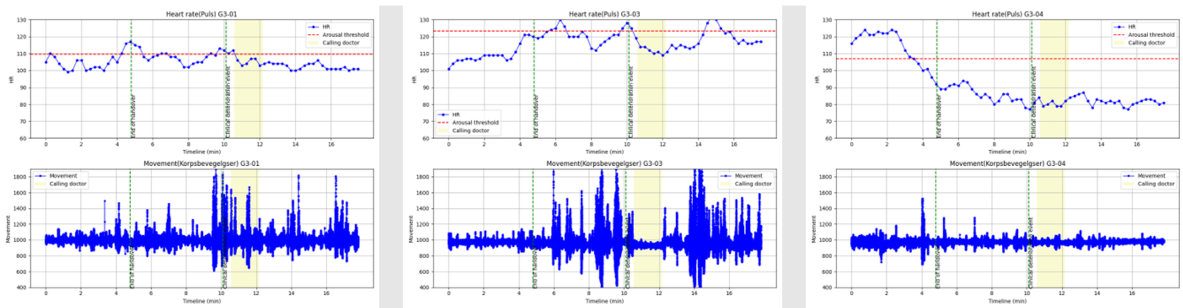


**Figure 1. MMLA-based feedback elaborated with the elicitation process**

Preliminary findings show that students, initially surprised by stress indicator visualizations, recognize their stress level from the stress indicator during simulation. They value this feedback for self-awareness but showed concern about its continuous use without actionable steps for improvement. This highlights the potential and challenges of MMLA-based feedback systems in specific educational settings, yet its broader applicability lies in the adaptable elicitation process for various learning scenarios with expert input and, to some degree, expected student performance.

## REFERENCES

Ericsson, K. A. (2008). Deliberate Practice and Acquisition of Expert Performance: A General Overview. *Academic Emergency Medicine*, *15*(11), 988–994.

Munshi, C., & Deneen, C. C. (2018). Technology-Enhanced Feedback. In A. A. Lipnevich & J. K. Smith (Eds.), *The Cambridge Handbook of Instructional Feedback* (pp. 335–356). Cambridge University Press.

Ochoa, X. (2022). Multimodal Learning Analytics—Rationale, Process, Examples, and Direction. In C. Lang, G. Siemens, A. Wise, D. Gasevic, & A. Merceron (Eds.), *The Handbook of Learning Analytics* (2nd ed., pp. 54–65). SoLAR.

Wise, A. F. (2019). Learning Analytics: Using Data-Informed Decision-Making to Improve Teaching and Learning. In O. O. Adesope & A. G. Rud (Eds.), *Contemporary Technologies in Education: Maximizing Student Engagement, Motivation, and Learning* (pp. 119–143). Springer International Publishing.

# Evaluating Pre-service Teachers' Science Teaching Practices in a Virtual Reality Classroom Using a Learning Analytics Spectator View

**Mafor Penn[1], Umesh Ramnarain[1], Christo Van Der Westhuizen[1], Ayodele Ogegbo[1], Noluthando Mdlalose[1], Koketso Pila[1], Martin Hlosta[2], Ivan Moser[2], Per Bergamin[2]**
University of Johannesburg[1] Fernfachhochschule Schweiz (FFHS)[2]
mpenn@uj.ac.za

**ABSTRACT**: More recently, virtual reality VR has emerged as one of the more desired learning technologies in Science, Technology, Engineering and Mathematics (STEM) education as it has been seen to enhance conceptual understandings of abstract concepts and provide a safe space for skill mastery. In this study, we evaluate pre-service teachers' science teaching practices when they teach microlessons in a VR classroom by using "spectator view" videos as the first part of a learning analytics dashboard feeder. Videos from the "spectator view" were analysed against PSTs' lesson plans to establish some of the gaps between planned and enacted microlesson instruction. Findings revealed that "spectator view" videos were an effective way to monitor and transform science teaching practices in the VR classroom and identify some gaps that needed reinforcement before PSTs teach actual learners. Teaching practices such as eliciting, questioning, probing, communication and group work for model labelling were well reinforced in the VR classroom, while investigations were poorly enacted. Future directions for this research are also discussed.

**Keywords**: Virtual Reality, Science Teaching Practices, Pre-Service Teachers Microlessons, Spectator View

## 1 BACKGROUND

One of the more prominent immersive learning technologies of the 21st century is virtual reality (VR). VR applications are computer-generated applications which immerse users into a completely virtual world closed out of their physical environment (Parong & Mayer, 2018). VR in education has been used extensively to acquire skills in mining, medicine, aviation and other vocational sciences (Pellas et al., 2020). More recently, VR has been seen to enhance conceptual understandings of abstract concepts, improve skill mastery, stimulate interest in learning and promote self-directed learning. Adding a learning analytics platform in VR provides an observer with the opportunity to evaluate what happens in VR and take measures for improvement. The Society for Learning Analytics Research (SOLAR) defines LA as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (http://www.solaresearch.org/about/).

In this study, a VR classroom application was developed by a joint Swiss-South African team and piloted with South African pre-service teachers (PSTs) training to teach biology, physics and chemistry at the high school level. The VR classroom was created for microlesson presentations and embedded with spectator cameras that feed into a learning analytics (LA) dashboard. The main aim of the study was to provide a judgement-free environment for PSTs to practice and prepare for teaching in actual classrooms. With PSTs targeted to teach in poorly resourced schools in mind, the platform allows PSTs to enhance their skill mastery and ability to teach using different science teaching practices and three-dimensional (3D) models. The envisaged six science teaching practices include eliciting, probing, and developing students' thinking about science: Choosing and using representations, examples, and models of content / Choosing and using representations, examples, and models of science content and practices: Supporting students to construct scientific explanations and arguments: Leading whole class discussions of content / Leading discussions that integrate science disciplinary core ideas and scientific practices: Setting up and managing small-group work / Setting up and managing small-group investigations: Establishing norms and routines for classroom discourse and work that are central to the content / Developing norms for discourse and work that reflect the discipline of science (Kademian & Davis, 2019). The problems that led to the creation of this VR classroom were threefold. 1. Lessons from the pandemic showed that PSTs needed a safe space for remote skill mastery, which could be used in scenarios where physical resources are scarce or not present. 2. Expert training using the equipped VR classroom could offset some of the shortfalls experienced in under-resourced settings. 3. Though environments for remote practice seem to be crucial for research in teacher education, an even more impending issue is how engagements in virtual learning environments can be evaluated by teacher educators. This problem led to a research question for the investigation.

- What are the prominent science practices visible in PSTs' practice during micro-lessons in a VR classroom when compared to their planned action?

## 2 METHODS

A qualitative research strategy using a case study design was adopted for the study, to analyse four (n=4) primary spectator view videos and four (n=4) lesson plans from a population of N=30 PSTs who participated in the pilot study. Selected participants were given pseudonyms gk3, mk21, mp29 and nt56 and all the relevant ethical considerations were followed strictly to keep their data and identity anonymous. During the experience, a PST could assume the role of a teacher (using the Oculus (Meta) pro VR headset) while others assume the role of learners (wearing the Oculus (Meta) Quest 2 headset). The interactions would be typical to those in traditional classrooms and more due to the availability of 3D models which can be manipulated by participants. Figure 1 below shows a sample layout within the VR classroom.

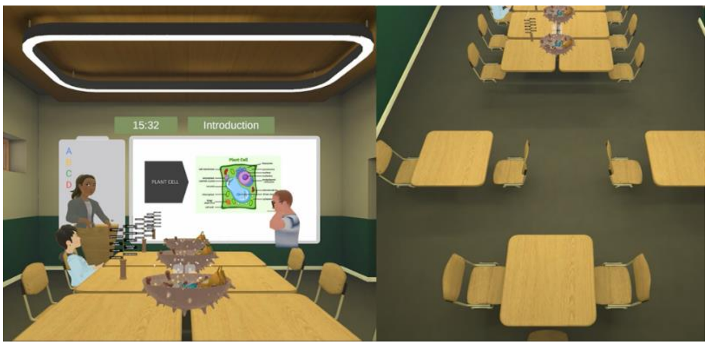



**Figure 1: Screenshot of VR classroom through the spectator's view**

Figure 1 shows a screenshot layout of the VR classroom where a PST assumes the role of a teacher while others assume the role of learners. A slide presentation can be made, and both teachers and learners can write on the whiteboard and be moved around in interactive sessions. 3D models seen on the virtual tables can also be manipulated. A deductive content analysis on spectator videos was conducted against PSTs' lesson plans to compare PSTs' planned actions and their actual practice in the VR classroom. Data were analysed deductively using a science practice rubric to capture science practices planned in the lesson plan and science practices enacted in the VR classroom from the LA spectator videos. PSTs' conversations and actions were coded with the rubric being the main coding scheme. This analysis assisted the participant researchers in evaluating gaps in PSTs' enacted practices to be addressed with the selected participants. Highlights of the role of the LA spectator view could also be isolated from the videos analysed

## 3 RESULTS

The primary results from the four participants indicated a common sequence of planning from the lesson plan analysis. PSTs planned to elicit prior knowledge in their lesson introduction, ask questions, probe for student reasoning, and use representations to teach the content. A quiz was also planned as a form of assessment in three of the four lessons. From the spectator video analysis, participants were seen to enact their planned actions, but much more than what was planned, other strategies, which involved learners working in groups to label 3D models and teacher-to-lead interactions, were supported. Eliciting, probing and questioning seemed to be the most prominent science teaching practice across participants while setting up and managing small-group investigations was the least enacted science teaching practice seen.

## 4 DISCUSSIONS AND CONCLUSION

From the findings in this research, the VR classroom gave PSTs the freedom to do more than was planned for their lessons as far as teaching in a VR classroom is concerned. The spectator view videos particularly provided a good platform for PSTs' and their educators to reflect on their challenges and improve their science teaching practices. Other challenges seen from the spectator view included lags in the feed that was coming from the VR classroom amidst other VR usability challenges. The study is still ongoing, and several limitations are associated with the networks and servers from which PSTs join the VR classroom. The researchers, therefore, recommend that similar remote teaching practice solutions be investigated in different contexts with the aim of enhancing PSTs' knowledge.

Acknowledgement

We are immensely grateful for the funding received from the Unity Social Impact and Meta Immersive Learning Programme

## REFERENCES

Kademian, S. M., & Davis, E. A. (2019). Planning and enacting investigation-based science discussions: Designing tools to support teacher knowledge for science teaching. In *Sensemaking in Elementary Science* (pp. 113-128). Routledge.

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology, 110*(6), 785-797. https://doi.org/10.1037/edu0000241

Pellas, N., Dengel, A., & Christopoulos, A. (2020). A scoping review of immersive virtual reality in STEM education. *IEEE Transactions on Learning Technologies, 13*(4), 748-761.

# Tackling the Alignment Problem: The Design of a Learning Analytics Dashboard for Teacher Inquiry

**Hong Shui[1], Xinran Zhu[2], Bodong Chen[2]**
University of Minnesota[1], University of Pennsylvania[2]
shui0003@umn.edu, xrzhu@upenn.edu, cbd@upenn.edu

**ABSTRACT**: Aligning learning analytics with learning design could offer opportunities for teacher inquiry and thereby support teachers to work as designers. However, in reality the alignment between learning analytics and learning design is often lacking. To mitigate the gap, we engaged three university teachers in co-designing a teacher-facing dashboard of social annotation activities under the guidance of the Activity-Centered Analysis and Design framework. By incorporating design elements as reference points of the analytics, the designed dashboard provides insights for teachers to reflect on and adjust their learning design and teaching practice. To investigate how the dashboard facilitated teacher inquiry, the dashboard was then implemented in four university courses in Fall 2023. This paper reports on findings from the implementations and teacher usage of the dashboard in diverse contexts.

**Keywords**: Learning design, learning analytics, dashboard, teacher inquiry

## 1 INTRODUCTION

The importance and value of the alignment between learning analytics and learning design have been highlighted in previous studies (e.g., Macfadyen et al., 2020; Schmitz et al., 2022). To guide the design and adoption of learning analytics, learning design lays a solid foundation for teachers (and other users) to inquire into learning and teaching processes (Law, 2017). When not properly aligned, data provided by learning analytics may not help teachers reflect on their learning design (Mangaroska & Giannakos, 2018), leaving the promise of learning analytics unfulfilled. Recent studies that aspire to address the misalignment have focused on refining methodologies and tools for co-designing learning analytics systems with stakeholders (e.g., Schmitz et al., 2022), engaging teachers in developing learning analytics dashboards (e.g., Kaliisa and Dolonen, 2023), and generating conceptual frameworks for the alignment (e.g., Law and Liang, 2020). However, several gaps still persist in prior literature: (1) the conceptual frameworks developed in most studies are overwhelmingly complicated or generic; (2) empirical studies remain insufficient, with a noticeable absence of reports on the impact of analytics-informed adjustments on learning design and teaching practice; and (3) learning analytics dashboards are mostly descriptive (of what things are) rather than conductive to action-taking (describing how things should be). To address these challenges, the study, informed by a learning design framework, involves teachers in co-designing a teacher-facing dashboard that seeks to align learning analytics with learning design and teacher inquiry. We report the findings of dashboard implementations in courses.

## 2 THEORETICAL FRAMEWORK

Among various frameworks of learning design, the Activity-Centered Analysis and Design (ACAD) framework has been incorporated into the study to guide the learning design and analytics on the dashboard (Goodyear et al., 2021). The framework conceptualizes two constructs: a design construct to document the learning design of teachers and an analytics construct for displaying analytics metrics with the reference to learning design. Its key elements – set design (e.g., resources and tools), social design (e.g., groups) and epistemic design (i.e., a sequence of tasks) – provide areas of consideration for eliciting teacher ideas during the dashboard design process.

The learning context used in this study is social annotation. Social annotation, assisted by web annotation tools, allows multiple users to annotate information in one shared document and anchor a discussion in the annotated information. It has been widely used by teachers to support students to achieve diverse learning goals. Social annotation offers a context where the design elements could be uniquely designed by the teacher.

By drawing on the ACAD framework and situating the dashboard design in the social annotation context, we asked the following research questions: RQ1) What design elements and inquiry would teachers choose to display on the dashboard for social annotation activities? RQ2) How does the dashboard support teachers to reflect and adjust their learning design and teaching practice?

# 3 METHODS

## 3.1 Context, Participants, and Phases

The study, which consisted of two phases, was situated in online/blended courses at a large public university in the United States. The co-design phase involved three researchers, three teachers who adopted a web annotation tool named *Hypothesis* for their classes, and two developers. This collaborative process unfolded in three stages conducted through Zoom meetings. In these meetings, teachers were engaged to outline their design elements using the ACAD framework, generate inquiry questions based on their respective designs, and co-design analytic measures. The result from the analysis of the meetings transcripts informed the dashboard design. The dashboard has two main components: *the design page* provides a space for the instructor to input design elements; *the analytics page* displays several analytics for social annotation activities. The design parameters are used as references to the analytics, delineating how things should be (Figure 1).

In the implementation phase, four teachers (two participated in the design) used the dashboard in their classes in fall 2023. In the first few weeks, researchers connected teachers online to prepare the dashboard utilization and sent out check-in emails every three weeks. Near the end of the semester, semi-structured interviews were conducted in Zoom. All interviews were recorded and transcribed.

## 3.2 Data Collection and Analysis

To answer RQ1, we examined the meeting transcripts of each teacher using descriptive coding and in vivo coding to extract their chosen design elements relating to social design and epistemic design, along with their corresponding inquiries (Miles et al., 2018). Then researchers developed themes by merging the similar ideas and preserving the unique ones. To address RQ2, we applied descriptive analysis to examine the log data and conducted content analysis on emails and interview transcripts.

# 4 FINDINGS

In response to RQ1, teachers defined specific participation roles within the social design component. In epistemic design, teachers devised various tasks, including annotating and replying, responding to prompting questions, tagging to classify annotations, and using sentence framing. When inquiring into social design, teachers focused on students' social interaction at individual and class level. Regarding the inquiry for epistemic design, teachers sought insights into students' engagement patterns. This included assessing whether students responded to prompted questions, analyzing annotations for signs of confusion, interest, connections, and conceptual understanding, evaluating the quality of annotations, and identifying topics that generated heated discussions.

In response to RQ2, initial findings were reported. By visualizing the analytics results, the dashboard supported teachers to understand students' interaction patterns and engagement in social annotation, providing actionable insights into the learning design and teaching practice. Informed by the dashboard, teachers grasped the enactment of the designed participation roles, identified students who did not participate, and filtered out students' confusions and heated discussion topics. Teachers enacted on the analytics by explaining the roles, reaching out to the absent students, and facilitating in class discussions.

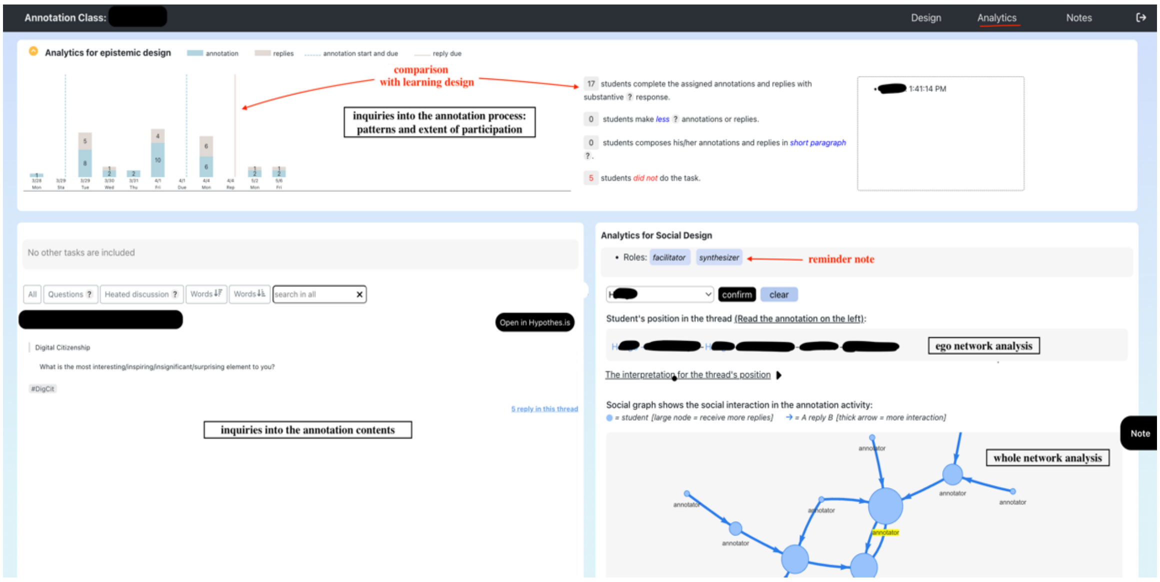



**Figure 1: Dashboard analytics page**

## REFERENCES


Goodyear, P., Carvalho, L., & Yeoman, P. (2021). Activity-Centered Analysis and Design (ACAD): Core purposes, distinctive qualities and current developments. *Educational Technology Research and Development, 69*(2), 445-464.

Kaliisa, R., & Dolonen, J. A. (2023). CADA: A teacher-facing learning analytics dashboard to foster teachers' awareness of students' participation and discourse patterns in online discussions. *Technology, Knowledge and Learning*, *28*(3), 937-958.

Law, N. (2017). Instructional design and learning design. In *The Sciences of Learning and Instructional Design* (pp. 186-201). Routledge.

Law, N., & Liang, L. (2020). A multilevel framework and method for learning analytics integrated learning design. *Journal of Learning Analytics*, *7*(3), 98-117.

Macfadyen, L. P., Lockyer, L., & Rienties, B. (2020). Learning design and learning analytics: Snapshot 2020. *Journal of Learning Analytics*, *7*(3), 6-12.

Schmitz, M., Scheffel, M., Bemelmans, R., & Drachsler, H. (2022). FoLA2—A Method for Co-creating Learning Analytics–Supported Learning Design. *Journal of Learning Analytics, 9*(2), 265–281.

# Assessing the Proficiency of Large Language Models in Automatic Feedback Generation: An Evaluation Study

**Wei Dai[1*], Yi-Shan Tsai[1], Jionghao Lin[2], Ahmad Aldino[1], Hua Jin[1], Tongguang Li[1], Dragan Gašević[1] and Guanliang Chen[1]**
[1]Monash University, [2]Carnegie Mellon University
[*]wei.dai1@monash.edu

**ABSTRACT**: Learning analytics (LA) exhibits profound potential in helping instructors with the laborious provision of feedback. Inspired by the recent advancements made by Generative Pre-trained Transformer (GPT) models, we conducted a study to examine the extent to which GPT models hold the potential to advance the existing knowledge of LA-supported feedback systems towards improving the efficacy of feedback provision. Therefore, our study explored the ability of two versions of GPT models -- i.e., GPT-3.5 (ChatGPT) and GPT-4 -- to generate assessment feedback on students' writing assessment tasks. We compared the feedback generated by GPT models (namely GPT-3.5 and GPT-4) with the feedback provided by human instructors in terms of effectiveness (content containing effective feedback component). Results showed that GPT-4 outperformed GPT-3.5 and human instructors in providing feedback containing information about effective feedback dimensions, including feeding-up, feeding-forward, process level, and self-regulation level.

**Keywords**: Learning Analytics, Feedback Generation, Generative Pre-Trained Transformer, Feedback Effectiveness

## 1 INTRODUCTION

The emergence of the field of LA exhibits profound potential in helping instructors with the provision of personalised and real-time feedback at scale (Arthars et al., 2019; Pardo et al., 2018). However, most existing LA-supported feedback systems still rely on instructors to assess students' performance and formulate personalised feedback, which poses a risk that students may receive feedback of sub-optimal quality due to instructors' limited capacity. We posit that Generative Pre-trained Transformer (GPT) models hold the potential to advance the existing knowledge of LA-supported feedback systems towards improving the efficacy of feedback. GPT models have shown their potential for feedback generation in the literature (MacNeil et al., 2022). However, limited studies have applied GPT models to generate assessment feedback on writing tasks with open-ended topics such as students' project proposals. This kind of writing assignment on open-ended tasks is common in higher education yet instructors often struggle to deliver comprehensive feedback for each student in large enrolment classes (Beckman et al., 2021). The current study compared the quality of feedback generated by GPT-3.5 with its advanced model, i.e., GPT-4, on writing tasks with open-ended topics. We assessed the effectiveness of feedback by using a well-known theoretical framework for feedback proposed by Hattie and Timperley (2007) and analysing the presence of effective feedback components in the feedback generated by GPT-3.5, GPT-4, and instructors. Specifically, this study was guided by the following research question (RQ): **To what extent does the GPT-generated feedback contain effective feedback components to guide student learning in comparison to human-produced feedback?**

## 2 METHODS

We retrieved the instructor-generated feedback from a postgraduate-level course teaching introductory data science skills. In this course, students were required to propose a data science project and submit a project writing proposal for marking. Instructors evaluated the submitted proposal and provided textual feedback for each student according to the following five aspects specified in the marking rubric: i) clear description of the goals of the project, ii) appropriateness of the topic to data science, iii) clear description of the business benefit, iv) novelty/creativity, and v) overall clarity of the report. After removing the student records without feedback, we finally obtained 103 students' proposal reports and the associated instructor-generated feedback.

We accessed GPT-3.5 and GPT-4 via the interface of ChatGPT developed by OpenAI. We designed the prompt for GPT-3.5 and GPT-4 as follows, "*Please give feedback on the following text in terms of a clear description of the goals of the project, appropriateness of the topic to data science, a clear description of the business benefits, novelty/creativity and overall clarity of the report. <INSERT THE TEXT OF A REPORT>*". For each student's proposal, we inserted the text of their proposal report into the prompt and submitted it to GPT-3.5 and GPT-4 to obtain generated feedback.

To answer the research question, we recruited two experts to annotate feedback by using the three dimensions (i.e., **feeding up**, **feeding back**, and **feeding forward**) and four levels (**task**, **process**, **self-regulation** and **self**) proposed in (Hattie and Timperley, 2007). After a pre-training session about annotation rules, each expert annotated 309 pieces of feedback in our dataset (103 pieces of instructor-generated feedback, 103 feedback generated by GPT-3.5, and 103 feedback by GPT-4).

## 3 RESULTS

**Table 1: The comparison of the distribution of seven effective feedback components between the feedback provided by the human instructor, generated by GPT-3.5 and GPT-4.**

| Components | Instructor | | GPT-3.5 | | GPT-4 | |
|---|---|---|---|---|---|---|
| | Quantity | Frequency | Quantity | Frequency | Quantity | Frequency |
| **Feeding up** | 6 | 5.83% | 0 | 0% | 97 | 94.17% |
| **Feeding back** | 101 | 98.06% | 102 | 99.03% | 103 | 100% |
| **Feeding forward** | 93 | 90.29% | 63 | 61.17% | 100 | 97.09% |
| **Task** | 103 | 100% | 103 | 100% | 103 | 100% |
| **Process** | 82 | 80% | 57 | 55% | 100 | 97.09% |
| **Self-regulation** | 11 | 11% | 0 | 0% | 18 | 17.48% |
| **Self** | 25 | 24% | 0 | 0% | 0 | 0% |

Table 1 indicated that GPT-4 was able to generate feedback containing effective components more consistently than human instructors, based on the prominent theoretical framework proposed by Hattie and Timperley (2007). Specifically, GPT-4 was superior to GPT-3.5 and even human instructors in providing feedback with more effective components, especially the **feeding-up** information, **process**-level and **self-regulation** level information. The high presence of **feeding-up** information in GPT-4 generated feedback boils down to the fact that GPT-4 explicitly included the assessment aspects in the marking rubric and connected them with student performance in the generated feedback to inform students of the attainment of learning goals. By contrast, GPT-3.5 and human instructors very rarely referred to the assessment aspects in the provided feedback. Our study also found that neither GPT-3.5 nor GPT-4 provided feedback at the **self** level which has been perceived as having a limited impact on enhancing learning gains (Hattie and Timperley, 2007).

## 4 IMPLICATIONS

According to the results of our study, GPT-4 yielded better performance in generating feedback that contains effective feedback elements. The observation presents an opportunity for instructors to leverage GPT models to improve the quality of their feedback by promoting GPT models to include more feeding up, process level and self-regulation level information. However, manually polishing feedback via ChatGPT interface for a large number of students can be time-consuming. Prompt design might be a challenging task for instructors without a background in engineering large language models. An alternative and ideal way is to encourage students to seek feedback from GPT models. By positioning the student as the central actor in the feedback process, the skills required to evaluate their own work, seek feedback from multiple sources, and use feedback to improve, commonly known as 'feedback literacy', can be developed (Carless and Boud, 2018). Given that LA can offer opportunities to trace students' engagement with feedback (Lang et al., 2017), the integration of GPT models into LA-supported feedback systems can close a full feedback loop (which is always at the core of LA (Clow, 2012). This loop aims not only to develop and enhance student feedback literacy by understanding students' sense-making and action-taking processes, but also encourage students to actively engage with their feedback (Carless and Boud, 2018).

## REFERENCES

Arthars, N., Dollinger, M., Vigentini, L., Liu, D. Y. T., Kondo, E., & King, D. M. (2019). Empowering Teachers to Personalize Learning Support: Case Studies of Teachers' Experiences Adopting a Student-and Teacher-Centered Learning Analytics Platform at Three Australian Universities. *Utilizing learning analytics to support study success*, 223-248.

Beckman, K., Apps, T., Bennett, S., Dalgarno, B., Kennedy, G., & Lockyer, L. (2021). Self-regulation in open-ended online assignment tasks: the importance of initial task interpretation and goal setting. *Studies in Higher Education*, *46*(4), 821-835.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315-1325.

Clow, D. (2012, April). The learning analytics cycle: closing the loop effectively. In Proceedings of *the 2nd international conference on learning analytics and knowledge* (pp. 134-138).

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

Lang, C., Siemens, G., Wise, A., & Gasevic, D. (Eds.). (2017). *Handbook of learning analytics* (p. 23). New York: SOLAR, Society for Learning Analytics and Research.

MacNeil, S., Tran, A., Mogil, D., Bernstein, S., Ross, E., & Huang, Z. (2022, August). Generating diverse code explanations using the gpt-3 large language model. In Proceedings of the 2022 ACM Conference on International Computing Education Research-Volume 2 (pp. 37-39).

Pardo, A., Bartimote, K., Buckingham Shum, S., Dawson, S., Gao, J., Gašević, D., ... & Vigentini, L. (2018). Ontask: Delivering data-informed, personalized learning support actions. *Journal of Learning Analytics 5*, *3*(2018), 235–249.

# Tracking the Evolution of Student Interactions With an LLM-Powered Tutor

**Kevin Gold, Shuang Geng**
Boston University, USA
klgold@bu.edu, sgeng@bu.edu

**ABSTRACT**: Student usage of an LLM-powered tutor to get homework help was tracked over the course of a semester in a university-level introductory data science course. For each homework assignment, the GPT-4 powered tutor was given the text of the homework problems and solutions in advance but was instructed to never reveal solutions directly, instead guiding the student to the correct answer through leading questions. Despite the free availability of ChatGPT, the majority of the class used the system. Anonymous logs were coded on seventeen dimensions of interaction. Evidence indicates that the students found the bot nearly as helpful as the human teaching assistants (TAs), and the bot was utilized more than the TAs' office hours. Some patterns of misuse, such as using the bot for convenient code checking, increased over the course of the semester.

**Keywords**: AI Tutoring, Large Language Models (LLMs), ChatGPT, Educational Technology, Learning Analytics, Qualitative Content Analysis, Higher Education, Programming Education.

## 1. INTRODUCTION

A looming challenge that large language models (LLMs) pose to the traditional homework model is that students could simply ask an LLM to do assignments for them Yet, LLMs have the potential to enable a degree of personal attention that would have been previously impossible at scale (Wu et al., 2023). This work explores the latter model of homework, where an AI tutor provides a steady supply of helpful hints without jumping too soon to the answer. While early work investigating the use of ChatGPT in education has rendered mixed results of varied performance across subject domains (Lo, 2023), we focus on evaluating the performance of using a GPT-4 API as an AI tutor in an introductory data science course at Boston University in Fall 2023. This tutor is given the homework assignment with solutions as part of its context, enabling convenient queries such as "How do I approach 1b," but is instructed to use the Socratic method to dole out guidance more frugally than ChatGPT. Three research questions about student's interactions with the tutor will be addressed: **1) How helpful can LLMs be in tutoring college students? 2) What are the typical bot failures and their trends over time? 3) What are the typical student behaviors and their trends over time?**

## 2. DATA & METHODOLOGY

The AI tutor provided a graphical user interface (GUI) for students to select an assignment via a radio button and enter a query. The student's query was augmented with a prefix: "For this query, answer with a single question that you haven't asked before that is meant to lead someone in the right direction, without directly answering the relevant homework question - unless the problem is solved completely, in which case, quit." The query was also augmented with the system information, "You are a helpful teaching assistant in a data science course. Your primary goal is to help the students learn. This is the homework the student was talking about: [homework text & solution]".

Two data sources are evaluated. The first one is a set of chatlogs of student-AI interactions. 802 sessions were recorded in total on eight homework assignments in a class of 127 students. We used a combination of deductive and inductive coding approaches to construct a set of initial interaction classifications based on chatlogs of the first two assignments. Our coding unit is each *problem* rather than *session* as students can ask about multiple homework problems in one session. Our final codebook includes 17 dimensions under four parent categories: A) Helpfulness of Advice; B) Bot Failures: Leak Correct Answer, Clear Error in Answer, Provide Irrelevant Answer, Bot Demands Extra Work, Fail to Point Back to Course Materials; C) Student Misuses: Select Wrong HW in GUI, Spam for Hints, Unclear Prompt, Search for Exploits; and D) Student Question Types: Debug Request, Review Code, Improve Style, Clarify Concept, Ask for Example, Recommend Resource, and General Hint Request. All dimensions are Boolean variables except A and B1, which are ordinal. We sampled 10 sessions from each of HW2 to HW7 for coding, with a total of 60 out of 802 sessions and 152 problems. (The easy HW1 mostly saw frivolous student behaviors, and HW8 was optional, so we excluded both.)

The second data source is a pair of surveys on students' experience with the AI tutor. 50 midterm survey forms and 65 end-semester survey forms were collected from 127 students.

## 3. RESULTS

**1) How helpful can LLMs be in tutoring college students?** Evidence indicates that most of the bot's answers are deemed very or moderately helpful by both the students (in surveys) and the independent coders with sufficient programming backgrounds (using chatlogs). Table 1 shows the percentage breakdowns before and after the midterm and across the data sources. In the end survey, students rank the AI tutor's helpfulness as equal to that of human tutors.

**Table 1: Helpfulness of the AI Tutor**

| | **Before Midterm** | | **After Midterm** | |
|---|---|---|---|---|
| | **Chatlogs** | **Midterm Survey** | **Chatlogs** | **End Survey** |
| **Helpfulness** (Very/Moderate/Not) | 86% / 9% / 5% | 29% / 62% / 9% | 84% / 8% / 8% | 42% / 49% / 9% |

While students find the AI tutor helpful in learning, they also utilize it more than the TA office hours. 69% of the respondents report that they have used the bot at least once, compared to only 46% who have visited TA office hours. Perceived ease of use, usefulness, and convenience have been identified as main factors of AI tutors' preferability (Malik et al., 2021).

**2) What are the typical bot failures and their trends over time?** Overall, we see a decreasing trend of bot failures in later assignments. Fisher's exact tests were performed across the five failure categories to discern significant trends, and statistically significant results were observed on B1-Leak Correct Answer (p-value=1.60e-04). Figure 1 (left) shows boxplots of the 4 levels of answer leaking across homework (4 levels: 0=no leak to 3=verbatim answer). The average level changes from 1.20 (near 1=mild leak) in HW2 to 0.34 in HW7 (almost no leak), showing that as problems became more multipart, the bot was less likely to leak the full problem.

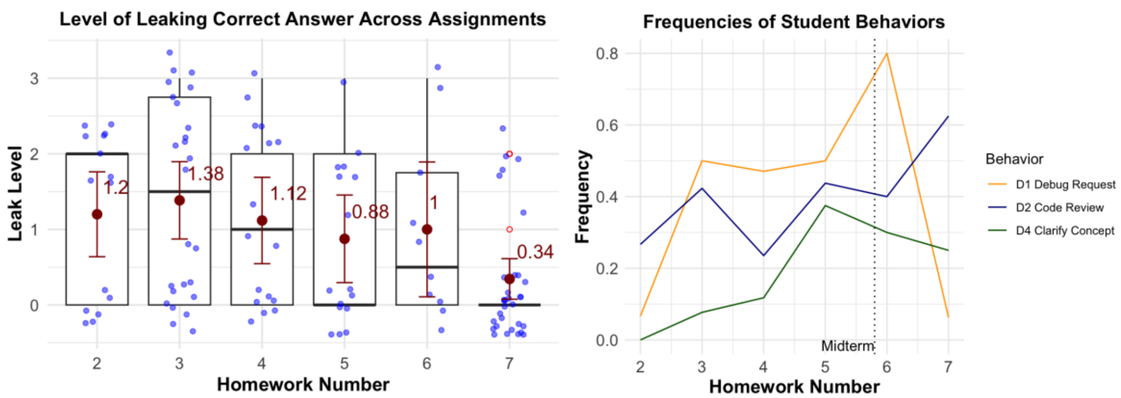



**Figure 1: Evolution of interactions. Left: Leak Answer Trend. Right: Freq. of Student Behaviors**

**3) What are the typical student behaviors and their trends over time?** Fisher's exact tests were performed across student misuse cases and question types to discern significant trends before and after the midterm. Statistically significant results were observed on D1-Debug Request (p=4.32e-05), D2-Review Code (p=4.19e-02), and D4-Clarify Concept (p=3.47e-02). The frequency of debug requests (asking the bot what is wrong with code in progress) went down from 0.46 to 0.08 (pre- to post-midterm), while that of code reviews (asking whether complete code is correct) went up from 0.36 to 0.58. One possible explanation is that the return of the graded in-class midterms generally convinced the class that they should not be overly reliant on the bot to debug programs (see Figure 1 right, showing frequencies by homework of the behaviors with statistically significant changes).

## 4. CONCLUSIONS

The evidence generally points to the AI tutor's advice being perceived as helpful by both students and coders evaluating the interactions. However, the evidence also suggests that misuse can rise over time as students learn to use the system in unintended ways, such as using it as a convenient answer checker or a "debugger" that is powerful enough to write the program one debugging step at a time (though in-class midterms serve as a deterrent to this behavior). It has been suggested that debugging is itself a useful activity for exercising critical thinking about a program and building the students' mental model of the program, the "notional machine" (Lowe, 2019). Further research will look at how to intercept or flag problematic interactions as they happen.

## REFERENCES

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, 13(4), 410. https://doi.org/10.3390/educsci13040410

Lowe, T. (2019). Debugging: The key to unlocking the mind of a novice programmer? *2019 IEEE Frontiers in Education Conference (FIE)*, 1-9. https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9028699

Malik, R., Shrama, A., Trivedi, S., & Mishra, R. (2021). Adoption of chatbots for learning among university students: Role of perceived convenience and enhanced performance. *International Journal of Emerging Technologies in Learning (iJET)*, 16(18), 200-212.

Wu, T.-T., Lee, H.-Y., Li, P.-H., Huang, C.-N., & Huang, Y.-M. (2024). Promoting Self-Regulation Progress and Knowledge Construction in Blended Learning via ChatGPT-Based Learning Aid. *Journal of Educational Computing Research*, 61(8), 3-31. https://doi.org/10.1177/07356331231191125

# "I Didn't Pass the Exam Because ...": Testing the Viability of Conceptual Features for Actionable Analytics in the Context of Competency Exam Failure Reflection

**Fanjie Li[1], Yeonji Jung[2], Alyssa Wise[1]**
[1]Vanderbilt University, [2]The University of Memphis
fanjie.li@vanderbilt.edu, yeonji.jung@memphis.edu, alyssa.wise@vanderbilt.edu

**ABSTRACT**: Taking an actionability-oriented approach, this study explored strategies for constructing predictive analytics for instructors to support students following competency exam failure. Our core approach builds on efforts in explainable modeling by making features in predictive models not only inspectable but also *interpretable* and *actionable*. We provide an example of this in the context of detecting dental students' failure attribution through analytics of their written reflections for failed competency exams. Through human-in-the-loop linguistic modeling, we built conceptual features describing students' perceived causes of failure to inform instructors in providing targeted support. We trained and evaluated a random forest classifier using the conceptual features and compared its performance with the classifier built on baseline n-grams features. LIME-based explanations of both models were generated for human interpretation. Results support the viability of conceptual features both in improving model performance and in enhancing model interpretability.

**Keywords**: Actionability, human-in-the-loop, conceptual features, explainable modeling, failure attribution.

## 1 INTRODUCTION

Being able to learn from failure and recover from difficult experiences is important for students' long-term success. Students' failure attribution, in particular whether they see the cause of their failure as within or beyond their control to change, has profound impact on how they perceive and react to the challenge, providing indications of whether the student is likely to recover independently or if external support is needed (Zimmerman & Moylan, 2009). Although students' attributional thinking is not always articulated and available to instructors, analytics of student reflections on failed experiences offers an opportunity for instructors to gain insights into how students see causes of their failure, identify who needs timely support, and recognize the specific areas in which students struggle. Using a corpus of dental students' written reflections on failed competency exams, this study explored strategies for constructing analytics of failure attribution with actionability to support recovery in mind.

## 2 STRATEGY 1: PRAGMATIC FRAMING OF OUTCOME CLASSES

Although the attribution theory (Weiner, 1985) provides nuanced dimensions for characterizing students' causal attribution (i.e., locus (internal/external), stability, controllability), fine-grained categorizations of failure attribution are difficult for instructors to interpret and act on within their busy daily routines. Zimmerman and Moylan (2009) have underscored the importance of the *controllability* dimension of attribution in self-regulated learning, suggesting that attributing failures to uncontrollable factors (e.g., external barriers, lack of ability) often discourages students' efforts for further improvement, whereas learners who attribute failures to controllable factors (e.g., use of strategies) are more likely to sustain motivation during setbacks and engage in further self-regulation. The alignment between self-regulated learning theory and dental education's focus on developing future dentists' ability for self-assessment and self-improvement (Driessen et al., 2005) motivated us to adopt an actionable framing of the outcome class (attribution for failure) that can directly inform instructors' decision-making as to *which students to support*: (1) **External/no attribution**: Students did not make attributions or attribute their failure to external factors only *(need timely support)*; (2) **Internal attribution:** Students demonstrate self-evaluation of knowledge, effort, etc. but do not show clear intent or path to change *(need timely support)*; (3) **Internal, controllable attribution**: Besides reflecting on internal reasons, students also demonstrate reflections on strategy or plan/intention to change. Students of this type are more likely to self-regulate (*less need for support*).

# 3 STRATEGY 2: CONSTRUCTING CONCEPTUAL FEATURES

To further support instructor in identifying the *kind of support these students need*, we examined whether conceptual features describing students' perceived causes of failure can be constructed via human-in-the-loop linguistic modeling. Informed by our prior thematic analysis of specific reasons for failure, we extracted noun phrases and specific forms of verb phrases (e.g., negation + verb such as "not study") that often capture linguistic cues of causal attribution. These phrases were subsequently input into ChatGPT 3.5 to assist in finding commonly used phrases to describe each reason for failure. The prompts to ChatGPT took the general form of "From the list of phrases below, please find all phrases that can be used to describe [REASON]". Based on ChatGPT's responses, we summarized linguistic patterns for detecting whether students mentioned a reason in their reflections. These include: (1) **Exam difficulty or delivery** (e.g., presence of hard/difficult/tough); (2) **Course design** (e.g., "clinical experience" that describe clinical exposure in curriculum); (3) **School activities** (e.g., "scheduling conflict" that describe hectic school schedules); (4) **Luck** (the part-of-speech tag "*existential there"* was used to indicate descriptions of external conditions given that sentence subject has been found to indicate writer's locus of control (Rouhizadeh et al., 2018)); (5) **Lack of knowledge** (e.g., presence of understand, grasp, clear about), (6) **Ability** (e.g., presence of (un)able/(in)ability to); (7) **Mistakes** (e.g., presence of verb starting with "mis"); (8) **Efforts to prepare** (e.g., presence of study/prepare/review); (9) **Learning/exam strategies** (e.g., time management).

In addition to perceived reasons for failure, we also constructed linguistic patterns for detecting **plan/intention to change**. Specifically, we used the presence of "*need to/more/further*", "*will*", and *comparatives* (e.g., more) to capture students' use of future-focus language, and used "*should/could have('ve)* [e.g., paid more attention]" to capture students' use of intention-intensive language.

# 4 MODEL PERFORMANCE AND LIME-BASED MODEL EXPLANATIONS

We trained a random forest (RF) classifier using the conceptual features and a second baseline RF classifier for comparison using n-grams features. For both classifiers, we performed a five-fold cross-validation on the training set for hyperparameter tuning and evaluated its performance on a 10% hold-out test set. The conceptual feature model achieved better classification performance (AUC: 0.83, Kappa: 0.49, precision: 0.68, recall: 0.67) than the baseline n-grams model (AUC: 0.79, Kappa: 0.47, precision: 0.66, recall: 0.64). We generated and compared LIME-based model explanations of three test-set reflections randomly selected from each class. Figure 1 presents model explanations for baseline and conceptual feature models for one of these instances (external/no attribution class) as an illustration due to space limitations.

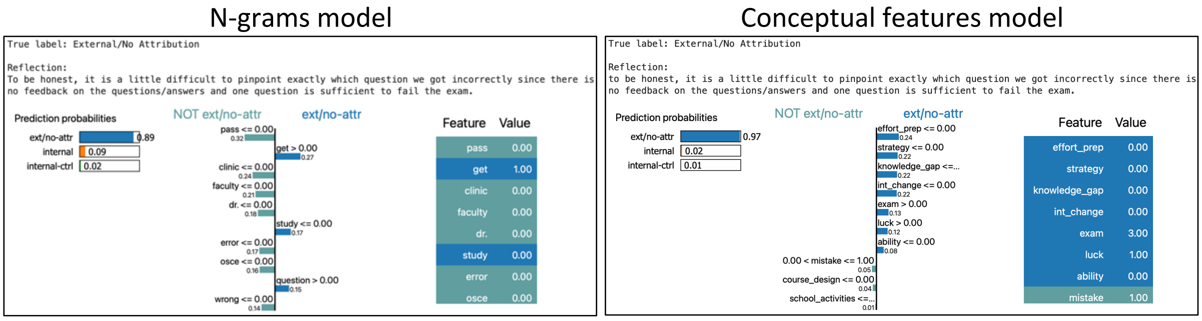



**Figure 1: LIME explanations for the instance randomly selected from external/no attribution class**

## 5 DISCUSSION

The results support the viability of conceptual features both in improving model performance and in enhancing model interpretability. Inspecting the model predictions for an instance from the *external/no attribution* class (Figure 1), both conceptual feature model and n-grams model correctly classified this instance. As aligned with our coding protocol, the reason that the conceptual feature model made the resulting prediction is that the extracted features do not indicate students' *intention to change* or attention to internal reasons (e.g., *lack of knowledge*, *efforts to prepare*, *strategy*, *ability*) (features <= 0) but does point to their reference to *exam difficulty* and *luck* (features > 0). Differential importance of such features across students (e.g. lack of knowledge versus efforts to prepare) is promising as a tool to point instructors towards different modes of support. It is noteworthy that the presence of the part-of-speech tag *"existential there"* (the matched pattern underlying the *luck* feature) rightly contributed to the prediction of the external/no attribution class. However, referring to the actual text of this reflection, *existential there* was actually used to describe the unavailability of feedback. This suggests that the "*luck"* feature may need to be redefined or split to avoid confusing instructors. For future work, we will further refine the conceptual features (e.g., the naming issue identified from the above observation) and evaluate the interpretability and actionability of the conceptual features with dental instructors and advisors.

## REFERENCES

Driessen, E. W., Van Tartwijk, J., Overeem, K., Vermunt, J. D., & Van Der Vleuten, C. P. (2005). Conditions for successful reflective use of portfolios in undergraduate medical education. *Medical Education*, 39(12), 1230–1235.

Rouhizadeh, M., Jaidka, K., Smith, L., Schwartz, H. A., Buffone, A., & Ungar, L. (2018). Identifying locus of control in social media language. In *Proceedings of EMNLP 2018* (pp. 1146–1152).

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, *92*(4), 548.

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In *Handbook of metacognition in education* (pp. 299–315). Routledge.

# Problem-Based Learning-Path Recommendations Through Integrating Knowledge Graphs and Large Language Models

**Hasan Abu-Rasheed[1], Rika Ikeda[2], Andrey Ferriyan[2], Christian Weber[1], Madjid Fathi[1], Keiko Okawa[2], Achmad Husni Thamrin[2]**
[1]University of Siegen, [2]Keio University
hasan.abu.rasheed@uni-siegen.de

**ABSTRACT**: Learning path recommendations are essential to acquire skill sets needed to solve real-life challenges. However, the main source of information that recommendation systems (RS) use to generate personalized paths is user data rather than the challenge that the user faces. In this research, we propose a problem-based approach to generate learning-path recommendations using knowledge graphs (KG) to connect learning materials, and large language models (LLM) for natural-language understanding and topic extraction. We construct a KG of courses and digital badges through human- and machine-extracted relations. Our RS analyzes a challenge written by the learner, extracts learning goals needed to solve that challenge, and then implements a Markov decision process (MDP) to select the optimal learning path. The learning path is then explained utilizing the KG and the LLM. We evaluate our KG relations in comparison to expert-defined tags. We also evaluate the recommendations and their explanations with a use-case approach. Our preliminary results show the ability of the proposed system to connect courses from different domains, recommend corresponding paths to the challenge requirements, and assign relevant explanations accordingly.

**Keywords**: Learning path, recommender systems, explainable recommendations, large language models (LLM), GPT-4, knowledge graphs.

## 1 INTRODUCTION

Solving real-life challenges requires multiple skills due to their complexity. Learning path RS have been developed in recent years to offer learners a sequence of learning materials that can achieve a learning goal (Nabizadeh et al., 2020). However, majority of RS use data about the learners themselves, e.g., in a user profile, to generate a personalized recommendation (Nabizadeh, 2020; Abu-Rashed, 2023). Analyzing real-life challenges that learners face and defining their learning goals are still difficult for automated recommendations. Therefore, it has been handled by human mentors. In scenarios where a challenge is temporal or dynamic, and when human support might not always be available, RS are required to 1) analyze the challenge, 2) break it down to its main learning requirements, and 3) generate a personalized learning path that enables the learner to solve the problem. These three tasks are, in fact, tasks of natural language understanding, topic extraction, and learning path generation. The recent development of large language models (LLMs) offers great potential to handle the first two tasks and support the RS in performing the third one more effectively (Zhao et al., 2023).

## 2 PROPOSED APPROACH

To accomplish the task of generating a learning path recommendation for a specific learning challenge, we developed an LLM-supported RS, which is based on a KG, to select, order, and recommend a set of courses that are relevant to the user-described challenge. In the proposed system, see Figure 1, users describe the challenges and learning needs in a free textual format. A GPT-4 LLM (OpenAI, 2023) then extracts the main topics from that description and phrase them as partial learning goals. We utilize prompt engineering to design a prompt context enriched with rules and contextual information from the KG, to mitigate the risk of model hallucination, irrelevant output, and too general responses.

The KG is a network of connected entities, which include in our case *study programs*, *courses*, and *digital badges* (Ikeda et al., 2023). To create the KG, we define a unique node type for each entity and connect the node using a hybrid approach, in which human-defined relations are complemented with a semantic relation extraction (RE) algorithm to enhance the connectivity of courses and badges. This offers more flexibility to navigate the KG and tailor the learning path recommendation to the problem at hand and the personal user profile. To generate the path recommendation, our RS: 1) considers the user profile as the starting point for any learning path. 2) identifies the group of partial goals extracted from the challenge description. 3) creates temporary relations between the partial goals. 4) Uses MDP to explore all paths and assign weights to each one, based on its compatibility with the challenge description and the user profile. The path with the highest weight is then recommended to the user.

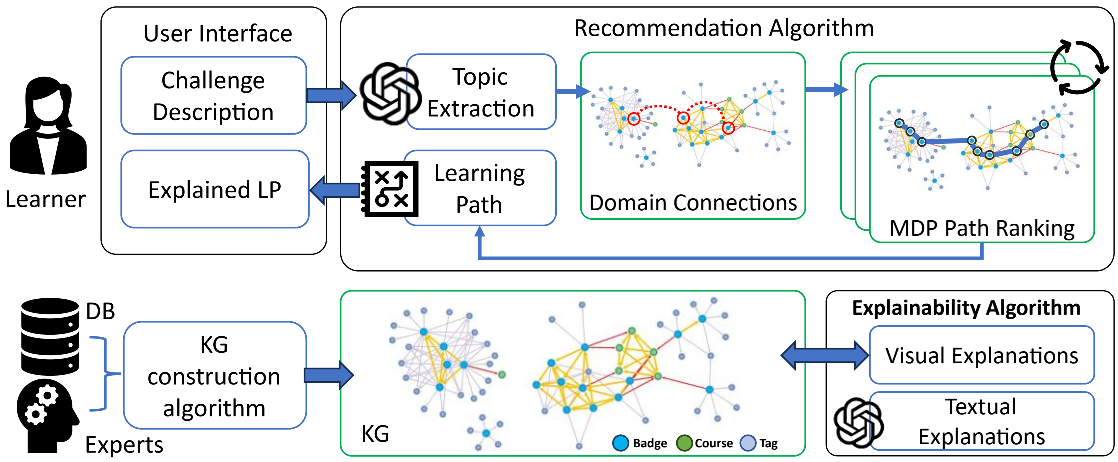



**Figure 1: Proposed approach for an explainable, graph-based learning-path recommendation**

The goal of creating temporary relations in the KG between the partial learning goals is to allow KG navigation algorithm to construct a learning path from multiple domains, which may not be connected by the human or the RE algorithm. An example in our case is *programming* and *water management*, which are semantically very different domains. Yet, both domains are necessary to solve, e.g., a challenge of analyzing geographic data in a flood crisis. We utilize an LLM-supported explainability approach to generate textual explanations of the learning path. Visual explanations are also generated from the KG and offered to the learner alongside the recommendation, supporting them in making an informed decision on the recommendation.

## 3 PRELIMINARY RESULTS

We evaluate the proposed approach on two levels: 1) evaluating the RE algorithm for KG creation, 2) evaluating the recommended paths and their explanations. The RE results are evaluated against human-defined relationships. We use the tags that creators defined in the course and badge metadata to indicate relations between elements that use the same tag. Semantic relations extracted by our algorithm from the textual descriptions of courses and badges covered 86.6% of the relations created through human-defined tags. To evaluate the explainable recommendations, we follow a use-case approach, in which complex challenges are defined and explainable learning paths are generated for them. An example challenge format is: "*A high level of pollution was discovered in a lake. I am organizing a team of volunteers to analyze the water supply data. For this task, I need to build a webpage to support volunteers' communication and write a Python program to automate the data analysis*". Recommended paths and explanations are then analyzed to evaluate three main measures: 1) correctness of topic extraction from the challenge, 2) acceptance of the recommended topics to solve the challenge, 3) acceptance of the path explanation. As a proof-of-concept, we used a set of 10 challenges, which were evaluated separately by 3 team members (one developer and two content experts) using a 1-5 Likert scale for each measure. Our preliminary result of the first prototype shows that the LLM-based challenge analysis scored 4.9/5 in identifying the main topics of the challenge. Recommending topics that are needed to solve the challenge scored an average of 4/5, while the path explanations scored 3.9/5. Our preliminary results also showed that the contextual information from the KG prevented the model from generating irrelevant texts. Textual explanations from the LLM were found to justify the connection to the challenge description. However, the LLM's output failed in explaining domain-specific terms or abbreviations when a clear definition was not explicitly provided for them in the data base. This limitation was also a part of the lower score for explanations, due to metadata sparsity. Encouraged by the proof-of-concept evaluation phase, we designed an extended evaluation strategy in the form of a user study, where we will survey a larger sample size of users who will interact with the system and evaluate the resulting recommendations and their corresponding explanations.

## 4 CONCLUTION AND FUTURE WORK

In this paper, we proposed an approach for generating explainable learning path recommendations, building on a KG, and utilizing LLMs for topic extraction from natural language. Our core contribution is generating learning path recommendations regarding a user-defined challenge alongside the user profile. This allows the path to be tailored specifically to a dynamic, urgent, or temporal problem the user faces. Our preliminary evaluation shows that the concept can achieve its promised tasks. A larger-scale user study, however, is required and will follow up this research for a deeper quantitative and qualitative evaluation of the proposed system.

## REFERENCES

Nabizadeh, A. H., Leal, J. P., Rafsanjani, H. N., & Shah, R. R. (2020). Learning path personalization and recommendation methods: A survey of the state-of-the-art. Expert Systems with Applications, 159, 113596. https://doi.org/10.1016/j.eswa.2020.113596

Abu-Rasheed H, Weber C and Fathi M (2023) Context based learning: a survey of contextual indicators for personalized and adaptive learning recommendations – a pedagogical and technical perspective. Front. Educ. 8:1210968. doi: 10.3389/feduc.2023.1210968

OpenAI (2023). GPT-4 Technical Report. ArXiv, abs/2303.08774. https://arxiv.org/abs/2303.08774

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J. R. (2023). A survey of large language models. arXiv preprint arXiv:2303.18223. https://arxiv.org/abs/2303.18223

Ikeda, R., Ferriyan, A., Okawa, K., & Thamrin, A. H. (2023). Enhancing Learning Mobility with a Community-Based Micro-Credential e-Portfolio Platform Service for Higher Education. In European Conference on Technology Enhanced Learning (pp. 566-572). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-42682-7_43

# Computer-Supported Code Generation Using Stepwise Coding

**Ayano Ohsaki**
Shinshu University
aohsaki@ohsaki-lab.net

**Daisuke Kaneko**
Hokusei Gakuen University
kaneko@hokusei.ac.jp

**ABSTRACT**: This study investigates the effectiveness of utilizing stepwise coding to support code generation for epistemic network analysis (ENA). To this end, we compared topics derived from stepwise coding with those extracted from raw interview data using the topic modeling technique. Through an interview with a teacher, this study applied the steps coding and theorization method (SCAT) to explore learning support technologies in music education. The findings indicate that topics identified through stepwise coding encompass meaningful words conducive to code generation, whereas topics derived from raw data contain noise and irrelevant terms. Furthermore, they indicate the potential of stepwise coding for facilitating computer-supported code generation. Future research endeavors will explore modeling the proposed approach and developing procedures for computer-supported code generation.

**Keywords**: Coding, Code generation, Stepwise coding, Quantitative ethnography, Steps for coding and theorization, Epistemic network analysis

## 1 INTRODUCTION

Exploring learning from both qualitative and quantitative perspectives has garnered substantial attention, with analytical tools like epistemic network analysis (ENA) expanding in popularity, offering visualization based on qualitative analysis (Shaffer, 2017). In analytical methodologies, code generation plays a pivotal role in data interpretation, with discussions emerging on the potential of automated approaches for code generation (Shaffer & Ruis, 2021). However, limited discourse exists on the process of generating inductive codes based on data. This study proposes using stepwise coding to facilitate code generation that incorporates analysts' understanding. Additionally, it compares the results of an automated approach between analyzing stepwise coding data and all raw data from teacher interviews on learning support technology in music education.

The study adopts the steps for coding and theorization (SCAT) as a stepwise coding approach, utilizing four steps: (1) extracting noteworthy words or phrases, (2) paraphrasing the extracted content, (3) accounting for paraphrased content with concepts outside of the text, and (4) composing themes considering the context (Otani, 2015). SCAT is a qualitative analysis method aimed at building a theory through two procedures: four-step coding and writing the storyline and theory (Otani, 2019). As illustrated in Table 1, this approach visualizes coding levels by delineating each step in corresponding columns. Although Kaneko and Ohsaki (2023) utilized SCAT to enhance the transparency of the connection between data and interpretations of ENA, no previous study has explored the potential of stepwise coding for computer-supported code generation. Consequently, we hypothesized that stepwise coding would be more conducive to generating code than raw data, such as interviews, as it would better capture analysts' understanding of actual phenomena. In this study, we employed the topic model for the automated approach, considering that there may be multiple topics within each data segment. The research question is: "How do topics estimated using data from stepwise coding differ from topics estimated using all raw data in an interview?"

## 2 METHODS

This study utilized interview data about the expectations of learning support technology in music education. The interviewer belonged to the development team of learning support technology, while the interviewee was a teacher associated with a brass band slated to test the support technology. The interviewer's objective was to uncover the teacher's perspective on music education and related music experience during the interview. The total number of utterance lines amounted to 361.

We utilized MATLAB (MathWorks, n.d.) to estimate topics from the data, employing the column labeled Step (4) and all utterances as raw data. Table 1 presents an example of the four-step coding method used for topic estimation. The second author, a qualitative researcher, conducted the coding process in Japanese. Subsequently, the first author translated the code content from Japanese to English. We applied a latent Dirichlet allocation (LDA) model (Blei et al., 2003) with seven topics to evaluate the standard topic modeling technique without special tool tuning.

**Table 1: Example of four-step coding.**

| No. | Speaker | Text (English) | Step (1) | Step (2) | Step (3) | Step (4) |
|---|---|---|---|---|---|---|
| 173 | Teacher | Right, right, right, I try not to do that, definitely. Of course, I would like to get a gold medal in competitions, but I don't want to win by neglecting the hearts and minds of students. | I would like to get a gold medal in competitions, but I don't want to win by neglecting the hearts and minds of students. | The desire for good grades, pressing the performers' emotions, desire for performance | The necessity of performance for rewards and the sacrifices necessary to achieve good grades | The trade-off between performance improvement and the players' emotions |
| 174 | Interviewer | I see. | | | | |
| 175 | Teacher | Yeah. So, when I feel that the sound is not in sync, I always ask [a question] like, "Do you know what your neighbor is playing?" It's not an issue about pitch. So, I want them to always keep in mind where they stand out and what the other person is doing; for example, "Oh, that's what I sound like when I play." | I want them to always keep in mind where they stand out and what the other person is doing; for example, "Oh, that's what I sound like when I play." | Understanding the piece, understanding the role of one's part, and thinking while playing | Teaching awareness of the music structure | Teaching performers to be constantly aware of the musical structure |

## 3 RESULTS AND FUTURE WORK

The results highlight the advantages of the proposed approach based on the four-step coding. Table 2 illustrates the top three clusters and frequent words within each cluster, derived from all utterance data and the data within the columns designated for Step (4) of the stepwise coding process. The analysis was conducted in Japanese, and the first author translated the frequent words in each topic from Japanese to English. Certain words enclosed in brackets in Table 2 were identified as frequent but could not be translated into English due to their lack of meaningful representation. Referring to the column labeled "From raw data" in Table 2, numerous words, including "Can" and "Do," appear as frequent words despite the removal of standard stop-words during preprocessing. This occurrence stems from the fact that the raw data consists of real conversations, encompassing cushion statements, interjections, and filler words typical of spoken language. Consequently, an automated approach from raw data necessitates more effective preprocessing and tuning to mitigate such occurrences. Conversely, in the column labeled "From stepwise coding" in Table 2, nearly all words carry meaning. Specific codes, such as "Instruction" from Topic 1 and "Performance" from Topic 2, could also be formulated. In further research, it is imperative to delve into the stepwise coding process to generate specific codes effectively.

This study was centered on comparing topics estimated from stepwise coding against those derived from all raw data to assess the potential of our proposed approach. Following the execution of stepwise coding, we identified meaningful words suitable as reference points for code generation, contrasting with the noisy topics derived from the raw data. This result underscores the validity of the proposed approach, even based on a singular example. This validity stems from its theoretical foundation in staged coding (Otani, 2019), wherein the steps facilitate the representation of data concepts. Theoretically, codes generated through the proposed approach encapsulate analysts' interpretations, thereby contributing to developing comprehensive ENA models. Future studies will compare models based on pure human code generation with our proposed approach. Additionally, the visualization of the relationship between topics and codes will be explored. Furthermore, we intend to investigate computer-supported code generation procedures, incorporating generative AI alongside the topic modeling approach. This integrated approach aims to enhance the fairness and effectiveness of analysis.

**Table 2: The differences of topics by data source.**

| Cluster | From raw data | From stepwise coding |
|---|---|---|
| Topic 1 | But, Right, Like, That, Instruction | Instruction, Sound, Due to |
| Topic 2 | Do, [iu], [teru], Understand, Then | Performance, Motivate, Novice, Objective |
| Topic 3 | Can, No, Say, Practice, [ii] | Practice, Agency |

## ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI (Grant Numbers 18K13238, 21K12178, and 23K11357) and by 2023 Hokusei Gakuen University Specific Research Grant.

## REFERENCES

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993–1022.

Kaneko, D., & Ohsaki, A. (2023). An approach to quantitative ethnography with SCAT: The use of qualitative. Data 61 Analysis Method, In Irgens, G. A., & Night, S. In the *Fifth International Conference on Quantitative Ethnography, Conference Proceedings* (pp. 61–62).

MathWorks. (n.d.). MATLAB. Retrieved December 15, 2023. https://www.mathworks.com/products/matlab.html

Otani, T. (2015). SCAT: steps for coding and theorization. Retrieved December 15, 2023. https://www.educa.nagoya-u.ac.jp/~otani/scat/index-e.html

Otani, T. (2019). *Paradigm and Design of Qualitative Study: From Research Methodology to SCAT*. The University of Nagoya Press, Aichi, Japan.

Shaffer, D. W. (2017). *Quantitative ethnography*. Cathcart Press, Madison, WI.

Shaffer, D. W., & Ruis, A. R. (2021). How we code. In A. R. Ruis & S. B. Lee (Eds.) *Advances in Quantitative Ethnography* (pp. 62–77). Springer.

# Designing a Learning Analytics Dashboard for Developing Online Teacher Productive Peer Talk

**Chao Yang, Gaowei Chen**
The University of Hong Kong
chaoyang@connect.hku.hk, gwchen@hku.hk

**ABSTRACT**: Although there is great potential for learning analytics dashboards (LADs) to enhance teacher learning and reflection on their classroom teaching, there has been limited investigation into the design of LADs to facilitate online teacher peer talk, which is essential for the development of their professional growth and instructional practices. Guided by the design-based research (DBR) approach, this paper elaborates on the process of designing a LAD to support the development of teachers' online peer talk that incorporates productive talk learning theories in three iterative cycles. Finally, an improved LAD version was provided, and the significance of the study was discussed.

**Keywords**: learning analytics dashboard, productive peer talk, online learning, teacher education

## 1 INTRODUCTION

Productive peer talk, which involves meaningful and constructive discussions among learners, promotes critical thinking and collaboration (Gillies, 2019; Lefstein et al., 2020). Researchers have investigated the use of Learning Analytics Dashboards (LADs) in supporting reflection, collaboration, and actionable insights for learners (Han et al., 2021; Susnjak, et al., 2022; Yoo & Jin, 2020). However, limited research has been conducted on the teacher aspect, especially on using LADs to analyze and support teachers' peer talk in their online professional practices. In addition, most LAD-related studies derive their data from system logs, overlooking the value of educator evaluations. To address the research gaps, this paper describes our in-progress work on developing a LAD that incorporates the theories of productive peer talk and utilizes teacher peer talk datasets toward their lesson videos, to better support teachers' learning and use of productive peer talk strategies.

## 2 LITERATURE REVIEW

### 2.1 Online Productive Peer Talk for Teachers

Productive peer talk is essential for promoting peer interaction, collaboration, and problem-solving (Hu & Chen, 2023). It is a dialogic process in which learners interpret data from multiple sources, and use it to enhance their work or methods (Carless, 2016). In an asynchronous way, online teacher productive peer talk is often conducted through reflective dialogue with peers based on the representations of their own teaching practices, such as lesson videos (van der Linden et al., 2022). When teachers watch, discuss, and give comments on each other's lesson videos, they gain professional experience from learning their peers' feedback, recognizing advantages and disadvantages, and places for enhancement. Researchers have identified a set of frameworks for analyzing teachers' productive peer talk. In this study, we adopted the framework of pedagogically productive talk (PPT, Lefstein et al., 2020), which aims to examine teachers' collaborative peer talk that is embedded in their day-to-day work to develop professional judgment. The PPT framework is comprised of six strategies for guiding productive peer talk: 1) problems of practice, 2) pedagogical reasoning, 3) representations of practice, 4) multivoiced, 5) generative orientations, and 6) support and critique.

### 2.2 LADs for Analyzing Collaborative Discourse

Analytics technologies have been increasingly used for generating data-driven insights, enhancing decision-making, and promoting innovation in the educational context (Susnjak, et al., 2022). One of the representative analytics tools is Learning Analytics Dashboards (LADs), which utilizes graphical representations that describe learners' academic and engagement levels to enhance self-reflection and encourage new insights (Yoo & Jin, 2020). Previous studies have shown the potential of LADs to analyze collaborative discourse, for example, Chen (2020) explored the effectiveness of a visual learning analytical tool, which visualizes classroom talk data to facilitate teacher reflection toward dialogic instruction. Suresh et al. (2021) investigated the use of TalkMoves, a LAD that displays visual representations of classroom talk moves to support mathematics teachers' reflection and teaching practice. Han et al. (2021) found that LADs can be used to support face-to-face collaborative argumentation between teacher and students. However, limited attention was given to developing teachers' online productive peer talk. This study addresses the research gap by designing a teacher-facing LAD that supports teachers' reflection on their productive peer talk and offers insights for teachers to make well-informed decisions on their teaching practice.

## 3 METHOD

The design of the LAD follows a design-based research (DBR) approach (Bell et al., 2013), as it allows researchers to examine factors that promote or hinder the intervention, thus bringing improvement in pedagogy and learning design (Anderson & Shattuck, 2012). By incorporating the theories of teachers' productive peer talk, we conducted three iterative cycles of DBR, and each cycle includes the pipeline of analysis, design, development, and implementation (Wang & Hannafin, 2005). Findings from each cycle were used for refinement of the design and development in the next cycle. Figure 1 depicts the improved version of the LAD design. Figure 1(A) provides an overview of the teachers' talk moves by time. Figure 1(B) shows the distribution of talk move types from different teachers. Figure 1(C) shows the interaction network among different teachers in a discussion group based on the number of talk posts.

## 4 SIGNIFICANCE OF THE STUDY

This study is expected to advance theory, methodology, and practice. On the one hand, this study addresses the lack of research in LADs that focuses on promoting productive peer talk from the teachers' perspective by incorporating principles of pedagogically productive talk into the LAD design. On the other hand, this study has the potential to offer implications for the design considerations of LAD for discourse analysis and teacher online learning environments. The findings of the study can be used to empower teachers with practical tools and strategies to enhance productive peer talk and instructional practices.

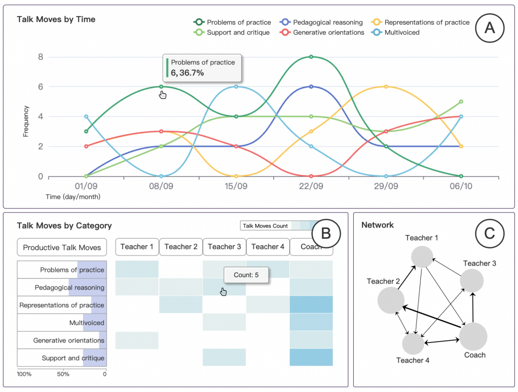



**Figure 1: The improved version of the LAD design**

## REFERENCES


Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher*, *41*(1), 16-25.

Bell, P., Hoadley, C. M., & Linn, M. C. (2013). Design-based research in education. In *Internet environments for science education* (pp. 101-114). Routledge.

Carless, D. (2016). Feedback as dialogue. *Encyclopedia of educational philosophy and theory*, 1-6.

Chen, G. (2020). A visual learning analytics (VLA) approach to video-based teacher professional development. *Computers & Education*, *144*, 103670.

Gillies, R. M. (2019). Promoting academically productive student dialogue during collaborative learning. *International Journal of Educational Research*, *97*, 200-209.

Han, J., Kim, K. H., Rhee, W., & Cho, Y. H. (2021). Learning analytics dashboards for adaptive support in face-to-face collaborative argumentation. *Computers & Education*, *163*, 104041.

Hu, L., & Chen, G. (2023). A systematic review and meta-analysis of productive peer talk moves. *Journal of Behavioral Education*, 1-33.

Lefstein, A., Vedder-Weiss, D., & Segal, A. (2020). Relocating research on teacher learning: Toward pedagogically productive talk. *Educational Researcher*, *49*(5), 360-368.

Suresh, A., Jacobs, J., Lai, V., Tan, C., Ward, W., Martin, J. H., & Sumner, T. (2021). Using transformers to provide teachers with personalized feedback on their classroom discourse: The TalkMoves application. *arXiv preprint arXiv:2105.07949*.

Susnjak, T., Ramaswami, G. S., & Mathrani, A. (2022). Learning analytics dashboard: A tool for providing actionable insights to learners. *International Journal of Educational Technology in Higher Education*, *19*(1), 12.

van der Linden, S., van der Meij, J., & McKenney, S. (2022). Teacher video coaching, from design features to student impacts: A systematic literature review. *Review of Educational Research*, *92*(1), 114-165.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, *53*(4), 5-23.

Yoo, M., & Jin, S. H. (2020). Development and evaluation of learning analytics dashboards to support online discussion activities. *Educational Technology & Society*, *23*(2), 1-18.


## ACKNOWLEDGEMENT

This work was supported by Hong Kong RGC, University Grants Committee (Grant No.: 17605221).

# Understanding the Dynamics of Student Workload and Difficulty: An Analysis of Weekly Perceptions of Higher Education Students

**Authors: Paul V. Sargent & Isabel Hilliger**
Pontificia Universidad Católica de Chile
pvsargent@uc.cl, ihillige@uc.cl

**ABSTRACT**: Study workload is widely recognized as one of the main concerns of higher education students. However, multiple issues arise when trying to effectively measure this construct, especially regarding autonomous work. In this study, we analyze data from a tool designed to measure students' perceived workload and difficulty, used by a total of 1,172 students at the Engineering Faculty of a Latin America University. Our findings show that one of the main factors related to the perception of higher workload is the workload distribution across course weeks. Also, perceived difficulty was highly correlated to academic workload, at both weekly and course level. Furthermore, we present a novel approach to identify periods of high intensity, referred to as "Peak Analysis". Finally, we discuss implications for instructional scaffolding that highlights study regularity and students' well-being, alongside possible lines of future research.

**Keywords**: Study Workload, Study Difficulty, Instructional Scaffolding, Study Regularity, Students Well-Being

## 1 BACKGROUND

One of the key variables that shape the experience of students in higher education is the management of academic workload (Smith, 2019). However, assessing students' workload addresses various issues. Measuring how much time students dedicate to their academic activities, especially on autonomous work, is a matter of high disparity which has drawn a lot of attention into the field in recent years. A traditional approach in higher education institutions is to assign credits to courses based on an expected number of hours students should invest weekly to fulfill their academic tasks. For example, Carnegie units are used in the U.S., while European institutions have their own Credit Transfer and Accumulation System (ECTS). Nonetheless, multiple evidence suggests that estimating student's workload from credit hours is insufficient. A recent study revealed that credit hours explain only a 6% of variance regarding course workload (Pardos, et al., 2023). Thus, estimating course workload effectively has gained fundamental relevance considering that credit hours is generally the only piece of official information students have at their disposal when choosing set of courses with a manageable amount of workload.

## 2 METHODOLOGY

This research presents the results of a Workload Survey, designed by Hilliger, et al. (2021) applied on the first academic period of 2023 (from March until early July). During this period, 1.178 students participated in this survey across 18 undergraduate courses. The survey was available online and gathered weekly reports of students' perceived workload and difficulty. Workload was measured as the number of hours students dedicated weekly to each of the course activities (which were defined by the teacher before the start of the academic period). Difficulty was measured on a Likert scale from 1 (Very Easy) to 5 (Very Hard), upon the course corresponding activities. As exclusion criteria, we first deleted responses with a workload above 168 (the total number of hours in a week). We used Tukey's fences method (Tukey, 1977) for outlier removal on each week responses outside the range (Q1 − 1, 5 * IQR; Q3 − 1, 5 * IQR). Finally, we set a minimum weekly response rate of 5%, and then accounted only for courses whose response rate was above 20%. We then estimated the average workload and difficulty, both at a course and weekly level, along with other measures of dispersion (Standard Deviation (SD) of weekly responses (within Week), SD of weekly workload (between Weeks), Ranges, etc.). Also, motivated by findings that highlight the importance of high intensity periods on students' perceived workload (Hilliger, et al., 2023), we created a novel approach for identifying these periods, to which we will refer as "Peak Analysis". This analysis involves identifying weeks where the workload is considerably higher than the course average, while also considering the context in which they occur. For a week to be classified as a peak, neither the preceding nor the following week can have a higher workload than the week in question. First and last academic weeks only consider the next and previous week, respectively. Peak weeks were categorized as follows: If one week workload is 25% above the course workload average, it's a Peak Type 1 (Low Intensity). If it's 50% above, then it's a Peak Type 2 (Medium Intensity). If it's 100% above, then it's a Peak Type 3 (High Intensity).

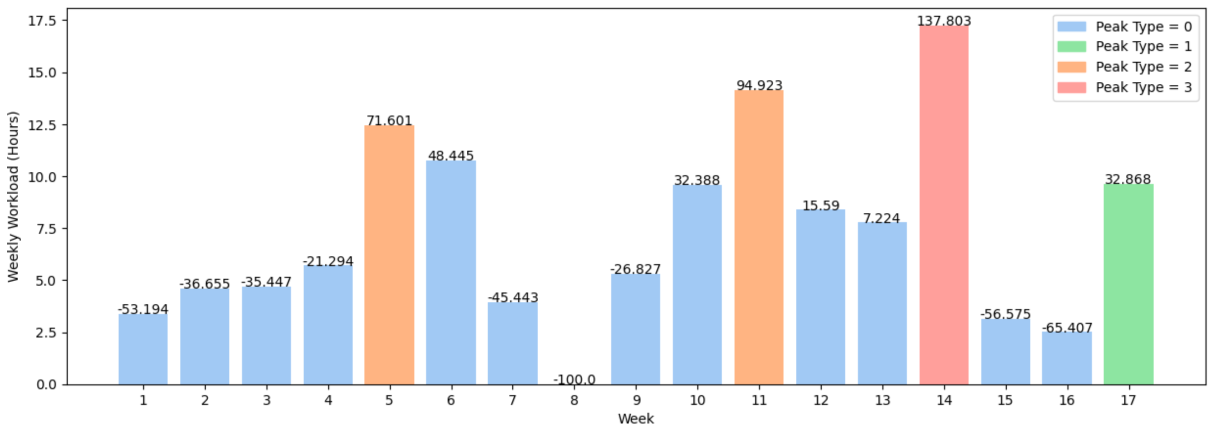


**Figure 1: Peak Type Categorization**

## 3 RESULTS AND DISCUSSION

Over a total of 78 constructed metrics, we conducted a Person's correlation analysis along with a statistical significance test for every pair of variables. We focused on identifying significant correlations for workload and difficulty, both at a course and week level. All the presented results have a p-value < 0.001. At course level, we found strong correlations between Course Workload and dispersion measures such as Course Workload Range (CWR) (r = 0.861) and SD of Weekly Workload (SDWW) (r = 0.821). In other words, courses with higher workload are strongly associated with greater variability in workload between course weeks. We found similar correlations, but to a lesser extent, between Course Difficulty and those variables (CWR, r = 0.53; SDWW, r = 0.506). This suggests that heterogeneity of workload distribution might play a key role on both student's difficulty and workload perceptions. We also found that Course Workload and Difficulty are strongly correlated (r = 0.746). Regarding week level metrics, we found that Weekly Workload is strongly correlated with SD of Weekly Workload Responses (r = 0.843). This indicates that weeks with a high-intensity workload are characterized by greater variability in the amount of time students allocate to the activities of that week. Weekly difficulty was also, as at course level, correlated with higher SD within week workload (r = 0.646). Furthermore, as expected, workload and difficulty were also correlated at a weekly level (r = 0.739). In relation to Peak Analysis, we found that both Course Difficulty and Workload were negatively correlated with the presence of Type 1 (Low Intensity) Peaks (r = -0.645 and r = -0.43, respectively). Conversely, Course Workload and Difficulty were positively correlated with Medium Intensity Peaks (r = 0.47 and r = 0.2, respectively). Additionally, Type 1 Peaks showed a negative correlation with both Type 2 (r = -0.35) and Type 3 Peaks (r = -0.24). Regarding Total Peaks (the sum of all Peak Types), Type 1 Peaks demonstrated the strongest correlation (r = 0.7), followed by Type 2 Peaks (r = 0.23), and lastly, Type 3 Peaks (r = 0.1). This indicates that more intense peaks are less frequent relative to the total number of peaks in a course, suggesting that further refinement in our categorization of Peaks might aid in capturing more precise nuances of workload intensity. Moreover, this could explain the lack of significant correlations between Type 3 Peaks and Course Workload and Difficulty. The above enhances our understanding of how variables related to workload distribution might shape students' perceptions of higher workload and difficulty, highlighting periods of high intensity, such as those revealed by our Peak Analysis.

## 4 CONCLUSIONS

These findings can serve as a crucial insight into making decisions that foster a manageable amount of workload for students. There is a wide corpus of research regarding study regularity based on LMS log data (Saqr, et a., 2022) but few studies have focused on how instructional scaffolding, especially course design, might be helping students to engage in course activities with a regular pattern. Considering that high intensity periods can negatively affect students' well-being (Hilliger, et al., 2023), we encourage educators to examine possible impairments between course workload and their credit hours. In such cases, workload distribution appears to be a key assessing point. Finally, a possible direction for future work is to further analyze the impact of these findings on other indicators, such as student's learning outcomes, academic performance or time management strategies.

## REFERENCES

Hilliger, I., Miranda, C., Schuit, G., & Pérez-Sanagustín, M. (2021). Self-tracking Time-On-Task: Web-Based Weekly Timesheets for Higher Education Students. In Technology-Enhanced Learning for a Free, Safe, and Sustainable World: 16th European Conference on Technology Enhanced Learning, EC-TEL 2021, Bolzano, Italy, September 20-24, 2021, Proceedings 16 (pp. 430-434). Springer International Publishing.

Pardos, Z. A., Borchers, C., & Yu, R. (2023). Credit hours is not enough: Explaining undergraduate perceptions of course workload using LMS records. The Internet and Higher Education, 56, 100882.

Saqr, M., Jovanovic, J., Viberg, O., & Gašević, D. (2022). Is there order in the mess? A single paper meta-analysis approach to identification of predictors of success in learning analytics. Studies in Higher Education, 47(12), 2370-2391.

Smith, A. P. (2019). Student workload, wellbeing and academic attainment. In Human Mental Workload: Models and Applications: Third International Symposium, H-WORKLOAD 2019, Rome, Italy, November 14–15, 2019, Proceedings 3 (pp. 35-47). Springer International Publishing.

Tukey, J. W. (1977). Exploratory data analysis (Vol. 2, pp. 131-160).

# A No-Code Environment for Implementing Human-Centered Learning Analytics Indicators

**Shoeb Joarder, Mohamed Amine Chatti, and Ao Sun**
Social Computing Group, University of Duisburg-Essen
{shoeb.joarder|mohamed.chatti}@uni-due.de and ao.sun.cn@stud.uni-due.de

**ABSTRACT**: Recent studies highlight the importance of involving stakeholders in Learning Analytics (LA), which has led to an emerging sub-field of LA known as Human-Centered Learning Analytics (HCLA). HCLA prioritizes human aspects in LA and integrates Human-Computer Interaction (HCI) methods to involve stakeholders throughout the LA process. While existing research showcases successful HCLA solutions with different stakeholders, these solutions focus on the participatory design of LA tools and platforms (macro design level) rather than the systematic design of the underlying indicators (micro design level). To fill this gap, in this paper, we present the conceptual details of the Indicator Editor in OpenLAP, a no-code environment that empowers end-users with no prior programming experience to steer the indicator implementation process based on their needs and goals.

## 1 USER SCENARIO

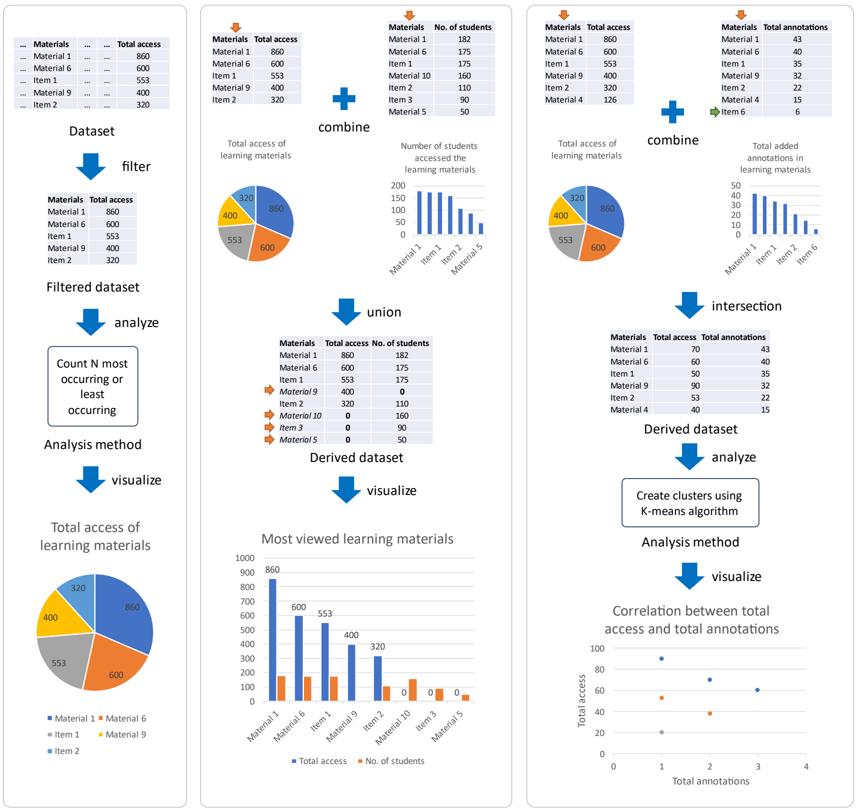
Jasmine is a professor at XYZ University who utilizes a MOOC platform to manage her course. She takes advantage of OpenLAP's personalized dashboard, which provides an overview of her courses using various indicators. On her dashboard, she has access to predefined indicators such as the participation rate of students in lectures, students' engagement in a discussion forum, and the progress of her students in assignments. Her particular interest lies in monitoring the number of accesses to her learning materials and the number of students accessing those materials. Upon searching through the available indicators in OpenLAP, she realizes that there isn't one that precisely exhibits this information. Consequently, she opened the *Indicator Editor* to create a customized indicator capable of performing statistical analysis, specifically counting. She calculates each learning material's total number of accesses for the first indicator. For the second indicator, she counts the number of times each student views the learning resources. She then adds these generated indicators to her personalized dashboard by embedding their respective code snippets.

## 2 INDICATOR EDITOR

The *Indicator Editor* in OpenLAP is responsible for providing end-users (e.g., teachers, students, researchers) who do not have prior programming experience, with an intuitive and interactive user interface (UI) that guides them throughout the entire indicator implementation process. In order to define an indicator, the users have to follow a four-step process: (1) *Dataset*: Explore the learning activities data available in the system and select an appropriate dataset, (2) *Filters*: Apply various filters to the dataset to make it more specific to the requirements, (3) *Analysis*: Select a suitable analytics method to analyze the filtered dataset, and (4) *Visualization*: Specify a visualization technique to visualize the analyzed dataset. There are three indicator types: *Basic*, *Composite*, and *Multi-level Analysis*.

# 3 BASIC INDICATOR

The *Basic Indicator* is a simple and easy-to-understand indicator type. Figure 1 (a) shows an example of such indicator, “Total access of learning materials”. This indicator creation process involves four key steps: Define a dataset, apply various filters, select an analytics method for analysis, and specify the visualization technique to visualize the indicator. Figure 2 (a) shows the steps of creating a *Basic Indicator* using the UI of the *Indicator Editor*.




(a) Basic Indicator (b) Composite Indicator (c) Multi-level Analysis Indicator

**Figure 1: Examples of the three types of indicators**

# 4 COMPOSITE INDICATOR

The *Composite Indicator* allows the combination of multiple *Basic Indicators* with different datasets and filters to create a more complex indicator. The primary condition to create this indicator is that all the selected *Basic Indicators* should use the same analytics method. The analysis results from each *Basic Indicator* are combined (union of the dataset) to form a cumulative analyzed dataset, which is then visualized using a visualization technique. Figure 1 (b) shows an example of how two *Basic Indicators*, namely “Total access of learning materials” and “Number of students accessed the learning materials” are combined to create a *Composite Indicator* “Most viewed learning materials”. Both *Basic Indicators* share the same analysis method, “Count N most occurring or least occurring items”. The datasets are combined, and the missing values in each column (marked with orange arrows) are filled with the value 0. The resulting indicator shows a group bar chart with “Materials” on the x-axis and “Total access” and “No. of students” on the y-axis. Figure 2 (b) shows the steps of creating a *Composite Indicator* using the UI of the *Indicator Editor*.

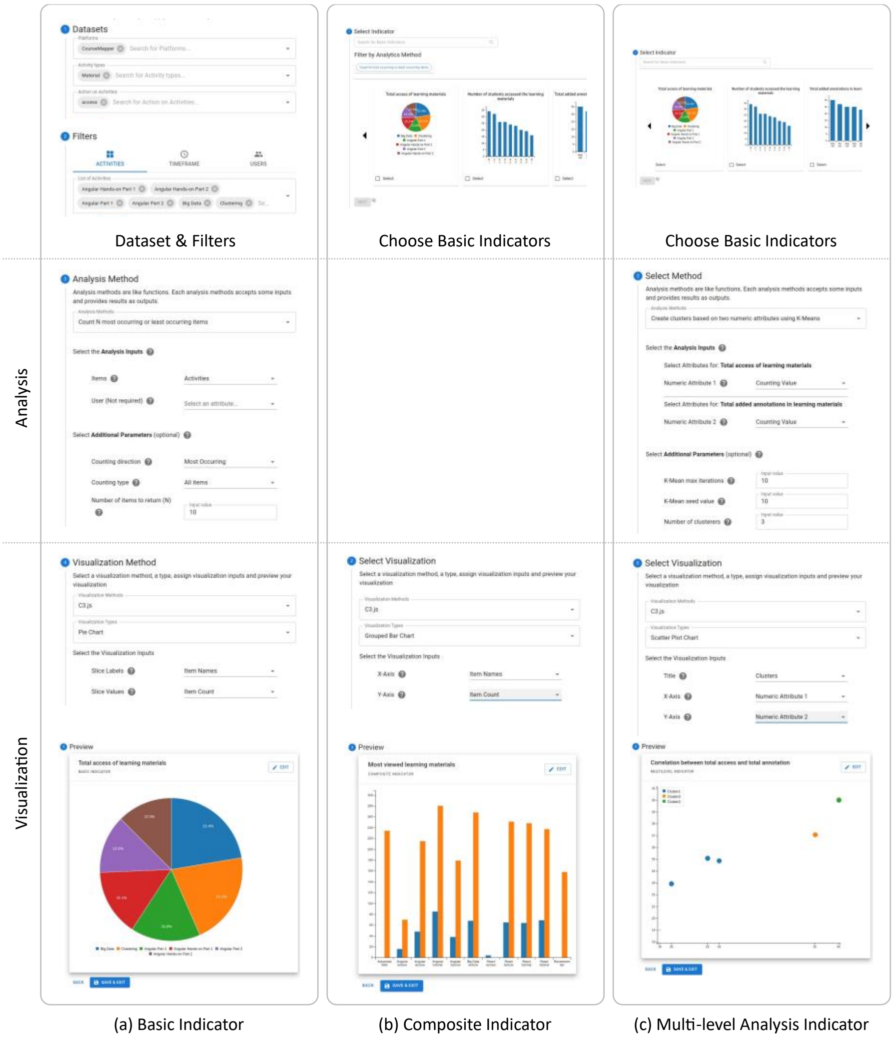



**Figure 2: UI of the Indicator Editor in OpenLAP**

# 5 MULTI-LEVEL ANALYSIS INDICATOR

The *Multi-level Analysis Indicator* allows the combination of multiple *Basic Indicators* with different datasets, filters, and analysis methods. The primary condition to create a *Multi-level Analysis Indicator* is that one of the data columns (attribute) must be common between the datasets of the *Basic Indicators*. The analysis results from each *Basic Indicator* are combined (intersection of the datasets) to form a cumulative dataset, which is used as the input of the second analysis method. The output of the second analysis method is then visualized using a visualization technique. Figure 1 (c) shows an example of how two *Basic Indicators*, namely "Total access of learning materials" and "Total added annotations in learning materials" are combined to create a *Multi-level Analysis Indicator*, "Correlation between total access and total annotations". The datasets of the two *Basic Indicators* have a common attribute, i.e., "Materials" (marked with orange arrows), and the intersection of these datasets removes data, which are not common between them (marked with a green arrow). The resulting dataset is used as the input of the analysis method to create clusters using the K-means algorithm, with three clusters (K=3). The resulting indicator shows a scatter plot with the clusters of data, "Total annotation" on the x-axis and "Total access" on the y-axis. Figure 2 (c) shows the steps of creating a *Multi-level Analysis Indicator* using the UI of the *Indicator Editor*.

# AI-ACS: An AI-based Course Articulation Coverage Score

**Sarah Goulart and Zachary A. Pardos**
UC Berkeley School of Education
sarahgoulart@berkeley.edu
pardos@berkeley.edu

**ABSTRACT**: As higher education within the United States relies on transfer pathways, evaluating institutions' effectiveness in aiding student transfers is crucial. This poster introduces the AI Articulation Coverage Score (AI-ACS), a novel metric designed to assess institutions' course articulation coverage for transfer students. By utilizing algorithmically estimated articulation as a baseline of potential articulations between sending and receiving institutions, AI-ACS aims to provide a precise evaluation of the completeness of course articulations between one institution and all the others in its system. This metric can serve diverse stakeholders, benefiting students in decision-making, aiding administrators in strengthening articulation strategies and guiding policymakers in system-level improvements.

**Keywords**: Student transfer, higher education, course articulation, institutional analytics

## 1 INTRODUCTION

In the 2015-16 academic year, almost half of bachelor's degree recipients in the United States from four-year institutions had previously attended a two-year public institution, highlighting substantial transfer activity (NSC Research Center, 2017). However, transfer students still face challenges when courses don't count towards their degrees, referred to as credit loss (Perez-Vergara & Orlowski, 2014). Articulation agreements, aimed at smoothing coursework transfers between institutions, play a crucial role in this landscape (Worsham et al., 2021). While previous efforts have explored Artificial Intelligence for creating new course-level articulations (Pardos et al., 2019), there's currently no metric for evaluating how well institutions articulate their courses to others'. This paper introduces the AI Articulations Coverage Score (AI-ACS), a proposed metric for assessing institutional articulations. Utilizing algorithmically estimated articulation as a baseline for potential connections between sending and receiving institutions, the AI-ACS aims to benefit various stakeholders—students, administrators, system heads, and policymakers—as part of an effort to enhance transfer processes.

## 2 ARTICULATION COVERAGE SCORE

This section elaborates on the development of AI-ACS. Initially, a simpler metric, without AI estimates, was explored to assess articulation coverage. A very basic approach would be to count the number of articulations associated with an institution. However, this naive method would exhibit bias toward institutions with large course catalogs. Another approach involved averaging the percentage of an institution's catalog size covered by course articulations within pairs, overlooking the connecting institution's size (i.e., ACS without AI). To refine this, we utilized an enhanced version of the AI-articulation estimation, inspired by Pardos et al. (2019), establishing a more reasonable upper bound on the articulation potential between two institutions. For calculating the average AI-ACS of a given institution (Inst. A), we examined all connecting institutions. Within each pair (Inst. A and Inst. B), we tallied existing course articulations. This count was then divided by all possible pairs of the AI-generated articulations, representing a percentage of all potential matching courses. This process was replicated for each institution pair, and the results were aggregated and divided by the number of institution pairs (n). In instances where an articulation pair involved multiple courses linked to one course, it was considered as a single articulation pair. For instance, Sociology 1A and Sociology 1B articulated to Sociology 100 would count as one articulation pair.

**Formula 1: AI Articulations Coverage Score (AI-ACS)**

$$AIACS(Inst\ A) = \frac{1}{n} \sum_{\substack{i=1 \\ i\ != Inst\ A}}^{n} \frac{\text{Existing Articulation Pairs(Inst A, Inst i)}}{\text{AI Estimated Articulated Pairs(Inst A, Inst i)}}$$

Formula 1 describes how the AI-ACS is calculated for an institution, offering an evaluation of the potential course articulations between paired institutions. This refined metric distinguishes itself from simpler assessment measures by leveraging machine learning, estimating the maximum potential eligible course articulations between institutions based on signals of course subject matter similarity. The utilization of machine learning allows for a more accurate representation of how close to its potential it is with respect to articulation coverage.

Following the formulation of AI-ACS, we calculated scores for each educational institution within the 64-campus State University of New York System (SUNY). Given its stature as the largest intersegmental higher education system, SUNY serves as a prime testbed for demonstrating AI-ACS. For this study we solely focused at pairs from a 2-year institutions to a 4-year institution. The calculation of these scores and their subsequent analysis is presented in Table 1, shedding light on the comprehensive articulation landscape within the SUNY System and its implications for transfer pathways and institutional collaborations.

**Table 1: Top 10 AI-ACS SUNY Institutions.**

| Institution Name | Overall AI-ACS | Std. Dev. AI-ACS | Max. AI-ACS | Min.AI-ACS |
|---|---|---|---|---|
| Columbia-Greene Community College | 0.261 | 0.204 | 0.771 | 0 |
| SUNY Cortland | 0.250 | 0.164 | 0.771 | 0 |
| SUNY at Fredonia | 0.244 | 0.127 | 0.414 | 0 |
| SUNY Oneonta | 0.210 | 0.130 | 0.588 | 0 |
| SUNY Buffalo State | 0.179 | 0.097 | 0.393 | 0 |
| SUNY College at Geneseo | 0.179 | 0.185 | 0.705 | 0 |
| SUNY College at Oswego | 0.155 | 0.097 | 0.435 | 0 |
| SUNY Adirondack | 0.150 | 0.101 | 0.358 | 0 |
| Dutchess Community College | 0.148 | 0.100 | 0.345 | 0 |
| Orange County Community College | 0.143 | 0.102 | 0.344 | 0 |

Table 1 showcases the top 10 institutions with the highest overall AI-ACS scores for 2-year to 4-year pairs of institutions. AI-ACS can be calculated for individual pairs of institutions. Table 1 also shows the standard deviations, minimums, and maximums representing AI-ACS for each specific pair. For instance, SUNY Oneonta has an overall AI-ACS of approximately .21 showing a 21% coverage of all potential course articulations to all possible connecting community college (CC). The minimum of 0 indicates that there is at least one CC without any articulations to SUNY Oneonta. Notably, all the top institutions have at least one institution with no course-level articulation, shedding light on potential coverage gaps for students.

The AI-ACS offers unique insights applicable to various stakeholders. All top articulated pairs of institutions have at least one institution with no course-level articulation, suggesting a need for system heads to enhance coverage across all institutions. The maximum value illustrates that, for all potential pairs of institutions, there is an opportunity for students to filter to a destination school or sending school of interest. For instance, Columbia-Green CC and SUNY Cortland have approximately 77% of all their potential courses articulated, indicating minimal credit loss for students transferring between them. The AI-ACS proves particularly beneficial for students exploring transfer options and administrators seeking to identify institutions in need of assistance in creating articulations.

The AI-ACS stands out for its ability to provide a balanced assessment of articulations, offering nuanced insights into an institution's interconnectedness with others. This metric contributes significantly to understanding how well an institution facilitates articulations across various pairs, guiding students towards informed transfer decisions.

The AI-ACS serves as a valuable tool for students and administrators, enabling a fair assessment of articulations and aiding in informed decision-making. While enhancing students' ability to evaluate course articulations and compare institutions, it also assists administrators in identifying gaps. Despite its comprehensive approach, the AI-ACS has limitations, including overlooking general transferred credit granted. Ongoing research aims to refine its accuracy by providing a more nuanced understanding of the number of institution pairs and differentiating between sending and receiving institutions. Despite these limitations, the AI-ACS stands as a valuable foundational metric for nuanced assessments and represents a step forward in addressing complex articulation challenges in higher education within the United States.

## REFERENCES

**NSC Research Center. (2017**, March 29). *Two-year contributions to four-year completions – 2017*. National Student Clearinghouse Research Center. https://nscresearchcenter.org/snapshotreport-twoyearcontributionfouryearcompletions26/

**Pardos, Z.A., Chau, H., Zhao, H. (2019*) Data-Assistive Course-to-Course Articulation Using Machine*** *Translation.* In Proceedings of the 6th ACM Conference on Learning @ Scale (L@S). ACM.

**Perez-Vergara, K. and Orlowski**, M. (2014). *A methodology for determining the transferability rate of courses.* Community College Journal of Research and Practice, 39(9), 793-807. https://doi.org/10.1080/10668926.2013.879376

**Worsham, R., Whatley**, M., & Loss, J. E. (2021). *Opportunity for all? the differential impacts of north carolina's revised comprehensive articulation agreement by race/ethnicity.* Education Policy Analysis Archives, 29(January - July), 28. https://doi.org/10.14507/epaa.29.5385

# An exploratory analysis of AI-supported essay writing and group peer assessment

**Kamila Misiejuk**
Center for the Science of Learning & Technology (SLATE), University of Bergen, Norway
kamila.misiejuk@uib.no,

**Aleksandra Pawelec, Tatiana Ershova**
University of Bergen, Norway
aleksandra.pawelec@student.uib.no, tatiana.ershova@uib.no

**Jarle Bastesen**
Kristiania University College, Norway
jarle.bastesen@kristiania.no

**ABSTRACT**: Peer assessment is often viewed as an approach to reducing an instructor's workload in big classes. In addition, it provides an opportunity for students to get to know the assessment criteria better and to develop their feedback skills. However, new challenges emerge with the spread of generative AI tools, such as ChatGPT, which may lead to students outsourcing all steps of peer assessment, from essay writing to giving feedback. This poster reports on a group peer assessment activity that involved collaborative essay writing. The data collection includes a survey of the use of AI, log data from the peer assessment platform and context data about the student. This preliminary research examined the variables associated with feedback patterns and student characteristics of feedback-receiving groups that used AI for their essay writing, as well as feedback-giving groups that suspected that the essay, they assessed, was written with the help of AI.

**Keywords**: Generative Artificial Intelligence, Peer Assessment, Collaborative Essay Writing

## 1 INTRODUCTION

There are conflicting interests regarding assessment from instructors' and students' perspectives. For instructors, assessment should spark reflection and support students' learning processes. For students, feedback is often perceived as an overall judgment of their work and the last step in their learning process. The emergence of easily available Artificial Intelligence (AI) tools, such as ChatGPT, complicates this situation even further. Students have access to many tools that can easily generate solutions to their assignments, such as essays. Peer Assessment (PA), a process where students evaluate each other's work, creates an interesting dynamic in this context. Instead of evaluating peers' work, students are faced with potentially assessing AI-generated texts and/or receiving peer feedback on the assignment that they produced using AI. Thus, they face similar issues and dilemmas that instructors are facing today.

This study focuses on exploring the effect of AI tools on group PA processes using learning analytics, and attempts to answer the following research questions: 1) How PA practices (both receiving and providing feedback) of students change depending on whether they have used AI tools to assist in their essay writing? 2) Does the use of AI tools in essay writing change student revision patterns? The preliminary results reported in this poster provide exploratory insights into the first research question through the analysis of the survey results on AI use by students as associated with the feedback characteristics, perception of peer feedback, student grades, and context information about the students.

## 2 DATA COLLECTION

Over 900 students participated in an undergraduate course at a Norwegian University in the Spring semester of 2023. After a couple of weeks of theory classes, the students were split into 2–4 students per group and tasked to work on an obligatory written assignment. Shortly after the course began, ChatGPT gained popularity. The use of AI tools was not prohibited by the instructor, but students were required to disclose its use in the text of their essays. An online platform, Eduflow (eduflow.com), was used to facilitate the PA activity. Eduflow was configured for each group to give anonymous feedback to two other groups based on 39-question-rubrics including a variety of free-text, binary and scale questions developed by the instructor. The groups received a 2-hour training in PA and were advised to use peer feedback to revise their essays. In addition, groups were given the opportunity to react to the feedback they received by giving comments and/or grading the usefulness of the feedback. Participation in the PA activity was obligatory, and the resubmitted essays were graded as the final exam (each group member received the same grade). After receiving the feedback on their own work, groups were asked to evaluate the feedback they received. At the end of the course, a short survey about the use of AI tools was conducted. The data collection has received ethical approval from the Norwegian Centre for Research Data. At the beginning of the PA activity, groups were asked for their consent to share their data for research purposes. This study adopted a strict consent policy, where all the data from non-consenting groups were removed.

## 3 DATA ANALYSIS & PRELIMINARY RESULTS

The collected data included the log data from the PA platform, final student grades, the initial and final drafts of their essays, and survey results. After data cleaning, the final dataset consisted of 193 unique groups (605 students) who consented, either gave or received feedback from a group that answered the AI survey or answered the survey themselves, and had complete data available, i.e. each member received a grade, and there is a draft and final version of the group assignment. Each feedback comment was coded by two researchers using the feedback model developed by Nelson & Schunn (2009) which focuses on feedback implementation, which is a well-established benchmark for effective feedback in the literature. Each code reached at least a moderate level of interrater reliability: *Summary* (κ=.67), *Praise* (κ=.92), *Identification* (κ=.70), *Explanation* (κ=.62), *Suggestion* (κ=.53), *Solution* (κ=.65), *Hedges* (κ=.56), *Mitigating praise* (κ=.82). Currently, the level of feedback implementation and student reflections are coded by a group of researchers.

A short AI survey was conducted after the PA activity and was answered by 74 groups. The results showed that most groups did not use the AI to write their essays (no=56, yes=18), most groups did not plan on using AI to write their final assignment (no=44, yes=20; NA=10), and most groups did not believe that the group that they were assessing used AI to complete the assignment (no=53, yes=16, NA=4). For the exploratory analysis, we used the survey results to construct two variables: 1) a boolean variable indicating, if a feedback-receiving group declared the use of AI (*used_AI*), and 2) a boolean variable indicating, if a feedback-giving group suspected that a feedback-receiving group used AI (*suspected_AI*). The variables used in the preliminary data analysis included: 1) feedback characteristics: the proportion of student answers to scale questions in the feedback rubrics, the sum of yes-answered boolean questions in the feedback rubric, the total length of feedback comments in characters, the total number of codes per evaluation (calculated separately for every code), 2) backward evaluation of feedback usefulness (on a scale from 1-5), and 3) context variables: previous experience with PA, group final grade, a binary variable indicating that a group plans to use AI to write the final draft and group size.

To examine the association between the variables *used_AI* and *suspected_AI* with other variables, exploratory data analysis was conducted using the Point-biserial correlation coefficient, which is best suited for a binary target variable (LeBlanc & Cox, 2017). The correlation results indicate that a feedback-receiving group that declared the use of AI in their essay writing was 1) more likely to use AI to write the final draft ($r_{pb}$=.6, $p$=.000), 2) was more likely to negatively self-assess the quality of their own essay ($r_{pb}$=-.19, $p$=.005), 3) more likely to receive a lower final grade ($r_{pb}$=-.15, $p$=.026), and 4) was less likely to receive an *Explanation* (a detailed explanation of a problem) in the feedback comment ($r_{pb}$=-.14, $p$=.038). The feedback-giving group that suspected the use of AI in the essay that they assessed was also more likely to declare that they are going to use AI in the final draft ($r_{pb}$=.45, $p$=.000), and was more likely to have group members with previous PA experience ($r_{pb}$=.13, $p$=.057).

## 4 DISCUSSIONS & CONCLUSIONS

Overall, the preliminary results on the actual or suspected use of AI were mostly not statistically significant and included weak or very weak correlations. This study captured the use of AI by early student adopters, who experimented with the newly emerging tools. The preliminary analysis does not show any distinct pattern in feedback-giving behaviour. The only feedback characteristic that was statistically significant showed a weak negative correlation between essays written with the help of AI and the likelihood of explanations in feedback comments. However, some interesting patterns could be discovered in the data. First, groups that either suspected the use of AI or used AI in their own essays were planning to use AI for the final essay writing. The answers to the open question in the AI survey about the reasons to use AI in the future reveal that most groups were planning to use it for inspiration or proofreading. Many groups also mentioned that AI will not be used to search for references. It indicates an early level of awareness of the potential opportunities and dangers of using AI tools by the students. The ongoing analysis of the feedback implementation patterns in the final essay will give more insights into how groups actually used AI to write the final draft. Second, the use of AI neither improved the final grade of the essay nor led to a positive self-assessment by students. Often, the use of AI is framed as a welcome shortcut for students to attain good grades, however, the influence of this practice on the actual performance or student perception of their own work needs more research in the future.

## REFERENCES

Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science*, *37*, 375-401.

LeBlanc, V., & Cox, M. (2017). Interpretation of the point-biserial correlation coefficient in the context of a school examination. *The Quantitative Methods for Psychology, 13*(1), 46-56.

# Experimental Interface for Multimodal and Large Language Model Based Explanations of Educational Recommender Systems

**Hasan Abu-Rasheed, Christian Weber, Madjid Fathi**
University of Siegen
hasan.abu.rasheed@uni-siegen.de, christian.weber@uni-siegen.de,
fathi@informatik.uni-siegen.de

**ABSTRACT**: In the age of artificial intelligence (AI), providing learners with suitable and sufficient explanations of AI-based recommendation algorithm's output becomes essential to enable them to make an informed decision about it. However, the rapid development of AI approaches for educational recommendations and their explainability is not accompanied by an equal level of evidence-based experimentation to evaluate the learning effect of those explanations. To address this issue, we propose an experimental web-based tool for evaluating multimodal and large language model (LLM) based explainability approaches. Our tool provides a comprehensive set of modular, interactive, and customizable explainability elements, which researchers and educators can utilize to study the role of individual and hybrid explainability methods. We design a two-stage evaluation of the proposed tool, with learners and with educators. Our preliminary results from the first stage show high acceptance of the tool's components, user-friendliness, and an induced motivation to use the explanations for exploring more information about the recommendation.

**Keywords**: Explainable AI (XAI), Recommender systems, Large language models (LLM), Chatbot, Multimodal explanations, OpenAI API, ChatGPT.

## 1 INTRODUCTION

Explaining educational recommendations to learners showed promising potential to enhance the learner's acceptance of the learning recommendations (Ooge et al., 2022). Explaining the recommendation is not only meant to clarify why certain content is recommended but also to support the learner's ability to make an informed decision about following the recommendation. This ability is greatly influenced by the type and volume of information learners receive from the explanations. Multiple types of explanations have been investigated in recent years. Hybrid and multimodal explanation approaches were found to increase the learner's satisfaction (Pecune et al., 20219) and engagement with the system (Tsai & Brusilovsky, 2019). With the new capabilities of LLMs, an emerging form of conversational explanations holds the potential to offer a better understanding of the recommendation by engaging the learner in a question-and-answer session about the recommended content. While the use of chatbots in education is not new, the use of LLM-powered chatbots in generating learning explanations is still under investigation, due to the great limitations of LLMs in a sensitive field like education. This reveals an essential and urgent need for creating tools that support researchers and educators in experimenting and evaluating different types of explanation modalities and approaches, as well as the different hybridization approaches amongst them. Our main contribution in this research is developing such an experimental tool, in the form of a modular and interactive interface, which allows the delivery of multimodal and chatbot-supported explanations.

## 2 EXPERIMENTAL INTERFACE OF RECOMMENDATION EXPLAINABILITY

We propose an interactive, experimental, web-based interface as a tool for delivering textual, visual, and chatbot-supported explanations of learning-recommendations. Our design is based on a pedagogical analysis of requirements expected from a learning-explanation. Pedagogy experts and educators were interviewed to determine these requirements, and how the tool's design can be tailored to visualize them to the learners. As a result, we focus on providing the means to explain: 1) the recommendation's connection to the learning goal, 2) relations amongst the recommended topics, 3) connection to the user-profile, and 4) connection to the teacher's original structure of the materials. Additionally, open-ended questions are supported through a chatbot. Our components, see Figure 1, are designed to enable experimenting with different combinations of explainability methods, by selecting which components are visible to the learner. We offer the following explainability components for experimentation: ***Textual explanations***: which are provided in a variable-length text area. By incorporating Markdown, this component allows for versatile text presentations. ***Tag-based explanation***: this is a more specific form of textual explanation but allows the user to click on the tags to reach further information using hyperlinks. ***Hierarchical structure***: which allows a clear overview of a recommendation's hierarchy, such as the one created by a human educator. It offers expandable elements and clickable titles with hyperlinking.

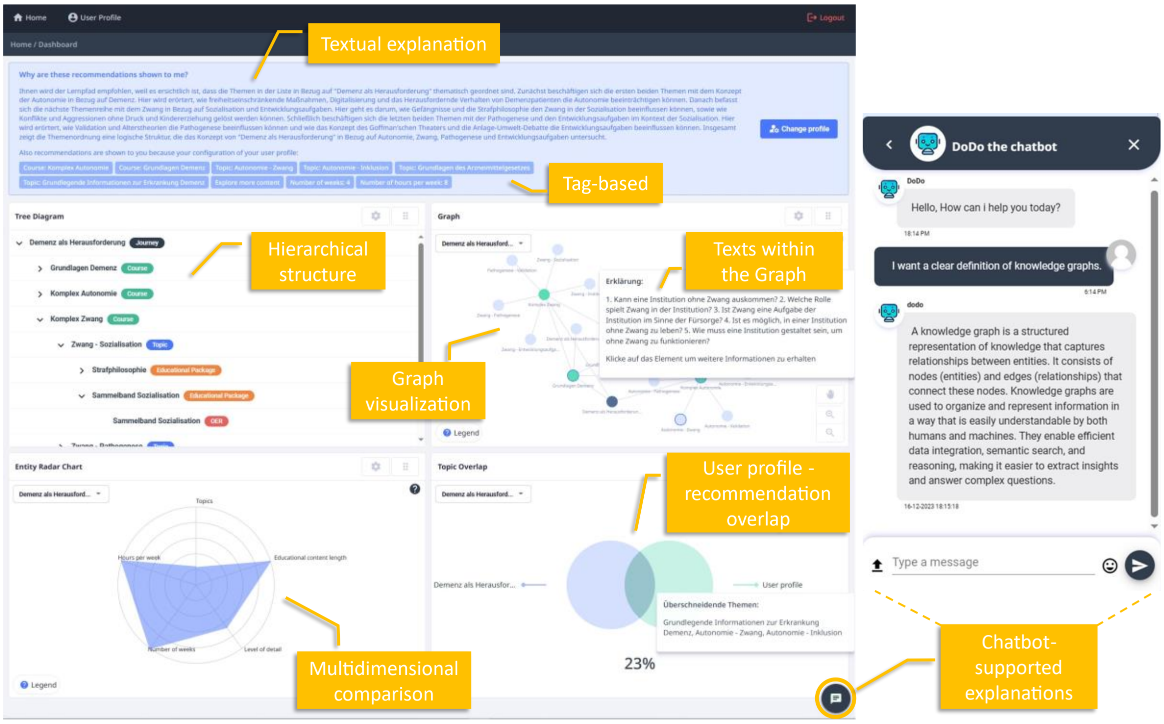



**Figure 1: Proposed interface with the complete view of explainability components**

***Graph-based visual explanations***: it provides interactive graphs for visual explanations, which can also show textual elements on the graph. ***Radar charts***: offering the potential to create multidimensional comparisons, to explain, e.g., coverage of the recommendations or overlap with preferences. ***Venn diagrams***: which provide a more focused overlap view, with a limited number of dimensions, but with interaction features including clicking, hover-over, and numeric or textual overlays. ***Chatbot-based explainability***: This component offers flexible support for chatbot use in explainability. We support LLMs using OpenAI API. Our support of LLMs is also achieved by the design of chat-messages to include contextual information and human defined rules, offering more control over the LLM's output, and thus reducing the risk of hallucinations and irrelevant responses. Our chatbot element enables researchers and educators to utilize the information on other elements of the interface, to enrich the LLM's prompt and context. It also provides a connection to the database, to acquire additional information about the recommended items. A multi-agent chatting support is also offered, allowing the user to chat with other users, mentors, or even different LLMs. This provides high flexibility for experimenting and evaluating chatbot roles in recommendation explainability.

# 3 EVALUATION AND RESULTS

We designed a two-stage evaluation strategy for the proposed tool. The first stage is to evaluate the tool's usability by the learners. The second stage is to evaluate the tool's usability by educators, and its features for accommodating singular and hybrid explainability approaches. We conducted the first evaluation stage with learners in the form of focus groups, in which learners were asked to perform a set of learning tasks on the interface. The learners then filled out a survey focusing on 1) their acceptance of the components, 2) the tool's user-friendliness, 3) the ease of retrieving information from the components, and 4) the tool's role in motivating the user to explore more about the recommendation. The second stage is planned in the next step of this ongoing research. In this stage, educators will test the tool with a set of pre-defined tasks. Qualitative interviews will be conducted to survey the educators' feedback on 1) the tool's usability, 2) most important features, 3) features that should be added to accommodate their experimentation requirements, and 4) most importantly, if the tools provide the sufficient level of abstraction for educators with a non-technical background to understand and utilize its features with minimal support from developers.

# 4 CONCLUSION

In this research, we propose an experimental tool that offers a comprehensive set of explainability components and features, to enable educators and researchers to test individual and hybrid explanation modalities and methods. Our tool is a web-based interface with interactive textual and visual modules. We introduce a chatbot element for conversational explanations, and offer thorough support for LLM-based chatbots, through accommodating OpenAI's API, prompt contextualization, rules, and connection to the databases. As ongoing research, we are continuing our evaluation plan with educators, and extending the features of the tool to include new visualization modules and more support for other LLM APIs, such as LlaMa and Gemini.

# REFERENCES

Ooge, J., Kato, S., & Verbert, K. (2022). Explaining Recommendations in E-Learning: Effects on Adolescents' Trust. In *27th International Conference on Intelligent User Interfaces* (pp. 93-105).

Pecune, F., Murali, S., Tsai, V., Matsuyama, Y., & Cassell, J. (2019). A model of social explanations for a conversational movie recommendation system. In Proceedings of the 7th International Conference on Human-Agent Interaction (pp. 135-143).

Tsai, C. H., & Brusilovsky, P. (2019). Evaluating visual explanations for similarity-based recommendations: User perception and performance. In Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization (pp. 22-30).

# Influencing Metacognitive Judgements with Perceived AI Annotations

**Warren Li, Christopher Brooks**
University of Michigan – Ann Arbor
liwarren@umich.edu, brooksch@umich.edu

**ABSTRACT:** In this study, we designed a tool to investigate the relationship between students' ability to render accurate judgements of learning (JOLs) when annotating their own work and comparing it with perceived AI-generated annotations. Our findings suggest that students rarely adjust their JOLs after seeing the AI annotations, indicative of a strong self-confirmation bias. The process of using the tool to self-annotate was found to enhance performance on a post-test. Emphasizing clear learning objectives and being transparent with the limitations of AI functions may improve the effectiveness of such tools as a way to provide quick feedback and mitigate hesitancy towards adoption.

**Keywords**: Metacognition, Judgements of Learning, Explainable AI, Self-Regulation

## 1 INTRODUCTION AND BACKGROUND

Metacognition is a crucial aspect of learning and involves the awareness and regulation of one's own cognitive processes. Monitoring one's own thinking, as well as changing behaviors based on this information is a crucially important skill for learners, which is increasingly captured using process-oriented methods such as log files (Azevedo, 2020). Constructs such as judgements of learning (JOL) fall in this category, which are assessments that people make about how well they have learned information (Pintrich et al., 2000). This is important as accurate metacognitive judgements have been shown to be helpful in educational activities such as time management and study planning.

A combination of human and automated feedback can lead to metacognitive development among learners. However, there is also a tendency of relying on rather than learning from AI (Darvishi et al., 2024), and machine-generated explanations come with their own limitations. One example is image or text classification, where linear interpretable model-agnostic explanations (LIME) highlight pixel groups or words to explain why a label was given. This supplement is not always used effectively and may subconsciously end up altering the user's beliefs to align with the explanation (Bauer et al., 2023). Interpreting visual representations of data is essential when making informed decisions, yet students often struggle. Deconstructing visualizations is a way to critically evaluate the effectiveness of the representation (Nolan & Perrett, 2016). This process also encourages metacognitive skills by prompting students to reflect on their understanding by articulating their own thoughts.

Consequently, in this study, we want to address and understand whether accuracy in metacognitive judgments is affected when students make their own annotations and are given AI annotations. We are also interested in exploring the case where learners first generate their own explanations, and if that affects their JOL and test accuracy. To answer these questions, we designed a tool to let students upload homework and reflect on their submission by making annotations, giving judgements of learning, and answering assessment questions on one platform. This enabled us to have students compare their annotations with ones supposedly created by an AI.

## 2 METHODS AND ANNOTATION TOOL

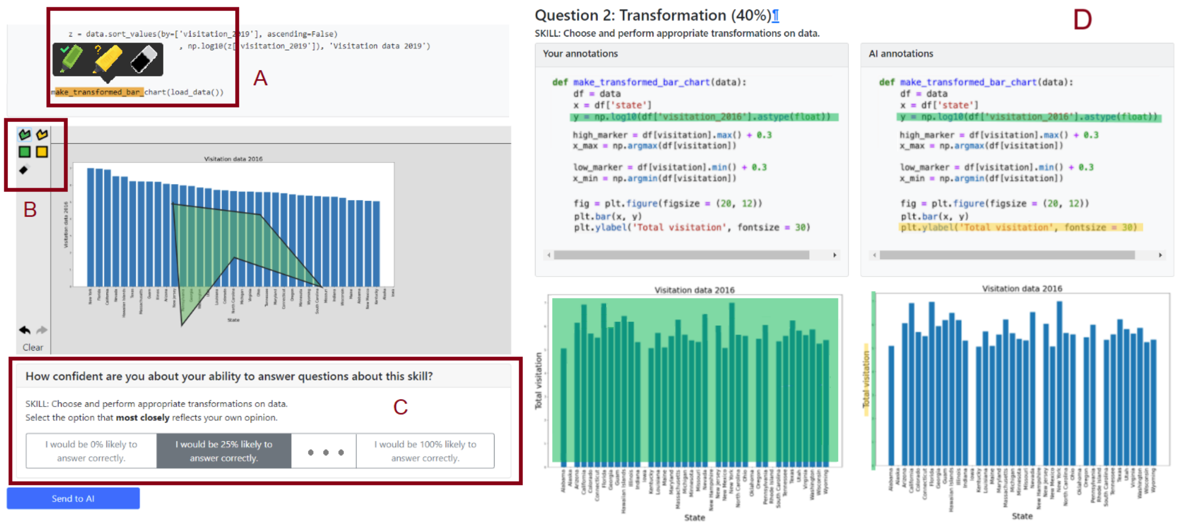



**Figure 1: A breakdown of the different components in the student-facing annotation tool**

To allow students to make annotations for later comparison, and leverage the benefits of self-reflection, our tool allowed users to color evidence: green markers demonstrated mastery of the given skill, while yellow indicated areas of doubt (Figure 1, Box A and B). This is similar to the LIME "superpixels" previously described. While visuals alone do not necessarily constitute a complete explanation, they do provide added value beyond a fixed numerical score. Whether they are more or less useful than text explanations depend on the existing knowledge of the user and the complexity of the task (Hopkins et al., 2020). However, focusing solely on visual annotations constrains time so that students can benefit from rapid feedback.

The method in which we elicit JOLs is based on word association experiments where participants were presented with a list of items or words and asked to estimate their future recall performance (Nelson & Dunlosky, 1991). This is less common in interdisciplinary contexts, but we applied these procedures by including a question about their confidence judgments (Figure 1, Box C) as well as a post-test to compare JOL ratings with actual performance.

Students could compare their own annotations side-by-side with the AI annotations to understand the influence of the given feedback (Figure 1, D). Afterwards, students were given the opportunity to change their JOLs before proceeding to the test items. In following human-computer interaction literature, we used the Wizard-of-Oz technique and leveraged a group of data science educators to generate the AI annotations and address our research questions in a timely manner; this was noted to students in the debrief.

## 3 PRELIMINARY FINDINGS

It was found that most students do not end up changing their JOLs. This may be because the feedback confirms existing beliefs, they began with strong convictions regarding their decision, or think that there is a misalignment between what they perceive to be important when declaring skill mastery and how concepts are assessed in practice. When one student was asked about why they did not make changes, they stated that, “A lot of it was a little bit tangential to what I was comfortable with.” Consequently, even though these questions were piloted with a few students and reviewed by the instructor, the relevance was not always clear to students. Thus, it may be important to ensure students understand why something is being taught and its importance, such as by directly referencing course material in AI explanations. Additional qualitative analysis from post-experiment interviews may provide more context into these behaviors.

**Table 1: Counts by whether students were required to make annotations. For each condition, this is split into scores received on the test items, as well as whether any change to JOLs were made.**

| Condition | 0 points | 1 point | 2 points | Changes | No Changes |
|---|---|---|---|---|---|
| Annotations | 7 | 19 | 14 | 13 | 71 |
| No Annotations | 15 | 14 | 8 | 16 | 62 |

There is a clear benefit to students in terms of having them first create their own annotations and then viewing the AI annotations side-by-side. For instance, given Table 1, while there is no conclusive link between those who choose to make JOL changes given annotations, there is a positive correlation ($r = 0.24$, $p = 0.035$) to suggest that those who received the annotation condition had better performance on the test items when compared to those who did not. This demonstrates the utility of self-reflections for learning and shows that the tool used in this study may be helpful for students to reflect on their coding or data visualization assignments.

## REFERENCES

Azevedo, R. (2020). Reflections on the field of metacognition: Issues, challenges, and opportunities. Metacognition and Learning, 15, 91-98.

Bauer, K., von Zahn, M., & Hinz, O. (2023). Expl(AI)ned: The impact of explainable artificial intelligence on users’ information processing. Information Systems Research.

Darvishi, A., Khosravi, H., Sadiq, S., Gašević, D., & Siemens, G. (2024). Impact of AI assistance on student agency. Computers & Education, 210.

Hopkins, A. K., Correll, M., & Satyanarayan, A. (2020, June). Visualint: Sketchy in situ annotations of chart construction errors. In Computer Graphics Forum (Vol. 39, No. 3, pp. 219-228).

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The “delayed-JOL effect”. Psychological Science, 2(4), 267-271.

Nolan, D., & Perrett, J. (2016). Teaching and learning data visualization: Ideas and assignments. The American Statistician, 70(3), 260-269.

Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). Assessing metacognition and self-regulated learning. Issues in the measurement of metacognition, 43-97.

# Use of Open Learner Models During Metacognitive Calibration

Warren Li, Christopher Brooks
University of Michigan – Ann Arbor
liwarren@umich.edu, brooksch@umich.edu

**ABSTRACT**: Open Learner Models (OLM) can help make the learning process more transparent. Being able to view information about their performance may also trigger self-reflection, which falls under the umbrella of metacognitive monitoring. While learning, utility of these tools may depend on the complexity of the task. People may fixate their gaze towards the OLM when they are looking for feedback, which might occur at certain intervals or on an as-needed basis if the item is difficult. As a result, we want to understand the relationship between OLMs can effectively foster more accurate JOLs by collecting both self-reported and behavioral data in the form of eye tracking. We hypothesize that using these sources in conjunction can provide insights into design decisions and use cases for OLMs.

**Keywords**: Metacognition, Triangulation, Open Learner Models, Eye Tracking, Trace Data

## 1 INTRODUCTION AND CONTEXT

Students who are better able to assess their own knowledge have been shown to attain higher academic performance (Fleming et al., 2016). Theoretically, a perfect calibration between one's perceived understanding and actual understanding, or judgement of learning (JOL), is desirable but difficult to achieve. JOLs are also subjective, and as with most self-reported measures, reactivity is hard to avoid; making a judgement will likely change the behavior that learners take under normal circumstances (Kit et al., 2018). Educational tools can help reduce discrepancies in self-monitoring and capture these processes.

Specifically, open learner models (OLMs) make some of the weights that factor into a prediction visible to learners. They allow students to inspect what a model thinks they know and compare it to their evaluation of their own knowledge, thereby supporting metacognitive reflection (Bull & Kay, 2016). Yet, making sense of the feedback provided by OLMs demands a degree of metacognitive competency since someone must assess the reliability of the help offered and realize when they should use external support; students may struggle with translating this data into effective action.

OLMs have evolved from simple, inspectable plots to ones with greater interactivity that allow learners to edit the model or present hierarchical structures such as concept maps. For fast decisions during an activity such as progress checking, simple and readily perceived interfaces may be preferred. Slower methodical thinking that encourages the user to reflect may benefit from more complex designs and interactive elements (Kay et al., 2020). Thus, the most effective options depend on context such as learner background and task complexity.

To better understand students' cognitive processes at a more granular level, other sources of data such as think-aloud protocols or activity logs can provide greater context (Winne, 2021). Trace data reflects the user's behavior as it occurs in real-time, and eye-tracking is one example that may prove useful to understand when and where individuals pay attention when learning new material. OLM interfaces consist of both text and graphs; eye-tracking can illuminate more precisely on which data representations students focus. For instance, whether a student refers to an OLM as a method of encouraging more reflective thinking when JOLs are elicited.

We were therefore interested in understanding how an OLM affects students' abilities to make accurate JOLs given recall tasks versus those that ask for application of knowledge. We also explore triangulation using self-reported judgements and eye-tracking data to understand how OLM use changes throughout the course of learning.

## 2 EXPERIMENTAL AND TOOL DESIGN

The study was conducted in a lab setting. In total, 80 participants were asked to answer questions about their prior music experience, complete 2 sets of learning tasks, and a final assessment. For each learning task, participants read a short passage regarding music in addition to answering 3 associated multiple-choice items. These items were all recall or application questions. Whether or not the OLM was displayed, as well as the ordering of question complexity was randomized.

The assessment consisted of 2 recall and 2 apply questions for each of 3 skills displayed in random order. Participant payout was a function of both their performance and JOL accuracy. The base payout was set at $8. For each additional correct answer, $0.50 was added. Participants were also asked about their likelihood of answering each question correctly and an additional bonus was paid out according to Table 1 in order to encourage participants to make a genuine attempt.

**Table 1: Payout is determined based on whether or not the item is answered correctly, and the confidence level chosen by the participant.**

| Option selected | If the item is answered **correctly** then the payout bonus is... | If the item is answered **incorrectly**, the payout bonus is... |
|---|---|---|
| Very Unlikely | $0.00 | $0.75 |
| Somewhat Unlikely | $0.25 | $0.50 |
| Somewhat Likely | $0.50 | $0.25 |
| Very Likely | $0.75 | $0.00 |

The OLM presented a total of 3 skills: instrument types, musical techniques, and musical traditions. Each question practiced fell under one of those categories. The design intentionally mimicked SQL-tutor (Mitrovich & Martin, 2007), an existing OLM used to teach learners about SQL, with slight changes in the visual design (Figure 1). This was chosen to ground the tool on an existing model that would be easy-to-use. As the learner answered questions, the progress bars relating to each skill filled in with a green color to represent correct understanding, and red for incorrect understanding.

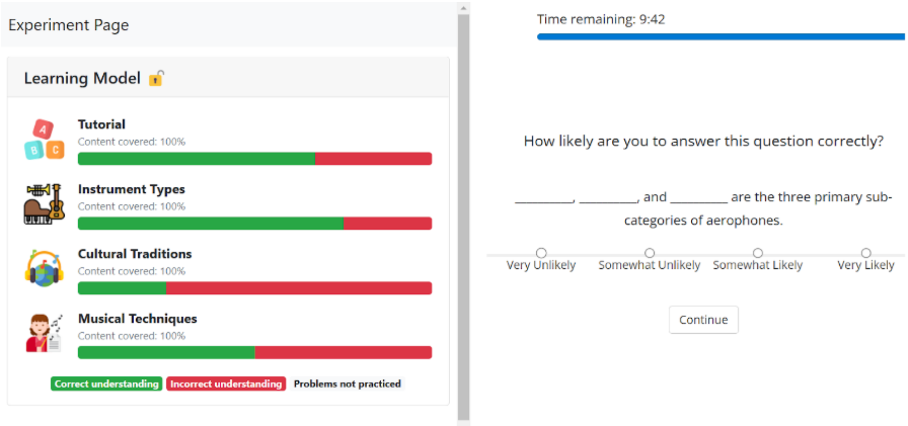



**Figure 1: The left-hand side shows the model which has been filled in after answering questions from the learning sessions. The right-hand side displays an example of a multiple-choice item and how the JOLs were elicited.**

## 3 CONCLUSION AND FUTURE WORK

By segmenting the screen into two halves, we can collect the duration spent looking at each half, as well as the number of times a learner's focus switches between left and right halves--this provides context into whether learners were paying attention to the OLM and contextualizes the situations when they do. For instance, whether participants take cursory glances only when the values changed after getting a question correct or incorrect, or even during periods between updates if they are thinking about whether a question falls under a particular skill. Lastly, the time spent on each question may serve as a proxy for item difficulty, and the total duration spent reading over each passage may provide insight into reading speed. Linking together these different measures such as self-report JOL and survey items along with eye tracking, can help us identify patterns in OLM usage with varying temporal resolution and guide future design considerations.

## REFERENCES

Bull, S., & Kay, J. (2016). SMILI☺: A framework for interfaces to learning data in open learner models, learning analytics and related fields. International Journal of Artificial Intelligence in Education, 26, 293-331.

Double, K. S., Birney, D. P., & Walker, S. A. (2018). A meta-analysis and systematic review of reactivity to judgements of learning. Memory, 26(6), 741-750.

Fleming, S. M., Massoni, S., Gajdos, T., & Vergnaud, J. C. (2016). Metacognition about the past and future: quantifying common and distinct influences on prospective and retrospective judgments of self-performance. Neuroscience of Consciousness, 2016(1), niw018.

Kay, J., Rus, V., Zapata-Rivera, D., & Durlach, P. (2020). Open learner model visualizations for contexts where learners think fast or slow. Design Recommendations for Intelligent Tutoring Systems, 8, 125-135.

Mitrovic, A., & Martin, B. (2007). Evaluating the effect of open student models on self-assessment. International Journal of Artificial Intelligence in Education, 17(2), 121-144.

Winne, P. H. (2021). Open learner models working in symbiosis with self-regulating learners: A research agenda. International Journal of Artificial Intelligence in Education, 31(3), 446-459.

# Exploring auto generated solution spaces of a serious game for introductory programing courses

**Paul Gamper***, **Jannik Decker, Ulrik Schroeder**
Learning Technologies Research Group RWTH Aachen University
*gamper@medien.rwth-aachen.de

**ABSTRACT**: In this poster, we present our approach to generate and explore solution spaces based on the data we collected during years of operating a serious game for introductory programming courses at our university. The idea of solution spaces is familiar, but relatively few research papers address the solution space in the context of novice programmers. Our solution spaces are automatically generated for all game levels from previous submissions and saved as an AST[1]-based graph. The invalid solution is added to the graph in the second step. Now, we can use the graph to generate adaptive feedback for the players. Visualization makes it possible to explore and compare solution spaces. We hope to find common errors and misconceptions and report them to the level designer and course teacher.

**Keywords**: Solution space, serious game, novice programmers, automated feedback, self-regulated learning

## 1 INTRODUCTION

There are many introductory programming courses at the RWTH Aachen University. The previous programming experience of the students is very heterogeneous. Various tools help the programming novices, including a self-developed serious game, Codescape. The game helps novices to take their first steps into programming in small steps and in a playful way in several introductory programming courses for Java, Python and C. The game has been used for six years, generating a large data set with over 2 million submissions from 7863 individuals. We want to visualize the collected data to explore the solution spaces and find hints about problems to improve the levels and change the order of the levels so it is easier for beginners. For the tutors, we want to point out possible misconceptions, which are visible through many incorrect submissions in the graph. Of course, the players should also benefit and get tips generated by the solution spaces. To address this issue, we are developing a tool to automatically create, store and update the solution space as a graph. The graph can be used for the on-the-fly generation of hints for the player. The graph can be visualized in the back-end of the game to help the creator of the level to identify some misconceptions on the one hand and analyze the user behaviour on the other hand.

## 2 RELATED WORK

In a small-scale study, Carbone et al. investigated how programming tasks can be improved due to poor programming habits of students in introductory programming courses. As a result, one of the suggestions was related to the solution space size of the task. The paper “Solution Spaces” Kasto et al. investigated this assumption and showed that the size of the solution space and the variety of

[1] AST – Abstract syntax tree, usualy a result of the syntax analysis phase of a compiler.

options for solving a task are related to the task difficulty. A similar serious game was used, and the target group was similar - programming beginners. The tasks can be more complex than the task designer might want them to be. However, this statement does not seem to apply equally to everyone, and it probably depends on experience. The solution spaces created by the teachers or tutors were often different to the solution spaces generated by the students. The solution spaces were created and analyzed manually. A promising automatic approach to generating solution spaces is made by Juha Helminen et al. for a tool to automatically generate the solution space. The players solve Parsons programming problems by moving the blocks with program code. The resulting solution spaces are used for investigations that uncover sub-problems of the tasks, but the method still needs to be adapted for real programming tasks on writing programs and not moving blocks.

## 3 SOLUTIONS SPACE VISUALIZATION AND HINT GENERATION

First, the submissions that are added to the solution space must be normalized. This includes: method and variable names renaming, such as removing comments, indentation and empty space. However, simplifying loops or conditions is separate from the normalization process. The way normalization is done may vary depending on the programming language and implementation. After the normalization is done, we use a directed and edge-weighted graph with a start and an end node to generate the solution space. The starting node of the graph always corresponds to the content of the first token (import statements in Java) and the last token corresponds to the end node of the graph(<EOF>). New entries are run through from the front and back and compared with the existing graph until a new node is found and added to the graph. The weight of the edges is then also updated. After the valid solutions, the invalid solutions are added in the same way.

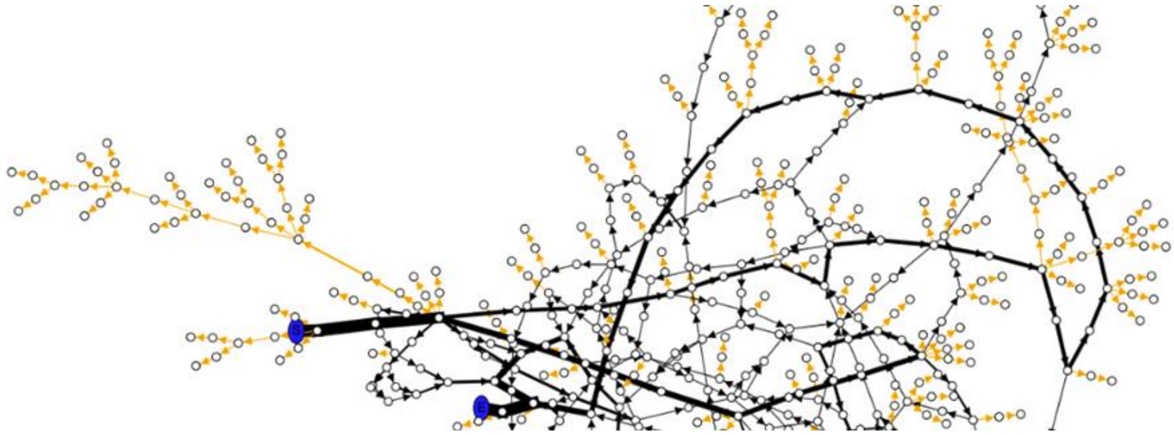


**Figure 1: Solution space of a level with minimum edge weight of 10. Black edges - solution space, yellow edges - invalid submissions**

For visualization (Fig. 1), the Java library GraphStream was used. The contents of the nodes represent the tokens. Multiple submissions of one player count only once. The edges are directed, and their thickness is determined by the ratio between the edge weight and the highest edge weight in the graph. To further improve the readability of the graph (e.g. 350.000 nodes), edge weights and node contents can be toggled on and off. Furthermore, the start and end nodes of the graph are colour-coded to make them easier to identify. The graph has two different types of edges: black edges representing the solution space and yellow edges representing deviations from the solution space.

To generate hints from the solution space, the process begins with normalizing the player's submission, followed by converting it into a token stream. Next, the solution space is compared to this stream. Once a deviation in the player's submission is detected, the path departs from the solution space at that point. At this exit point from the solution space, the nodes that should be taken instead of the deviation to re-enter the solution space can be determined. These nodes serve as "Next-Step-Hints" and can be communicated to the player. Furthermore, the player can also be informed about which token in their submission needs to be corrected. This process can also be carried out from the end of the graph.

## 4 CONCLUSION AND FUTURE WORK

We could automatically generate, store and use solutions spaces from user submissions. A visualization tool and a hint-generation prototype are developed. The token-stream approach was chosen to generate the solution spaces, which are generated from a grammar. This allows us to expand the approach to other programming languages. Whether the solution spaces of different languages can be compared still needs to be investigated.

We were also able to confirm the results of previous studies, namely that the size of the solution space correlates with the difficulty level or the novices. Some levels at the very beginning of the game and offering too large and open solution space will either be restricted or moved.

Future research will focus on how the solution space analysis and visualization can help the tutors of the lecture to understand the student's problems facing in a serious game at individual level. For this idea we need first to improve the visualization to show the single player progress as a subgraph. We hope also this technique will help us to discover common errors and misconception for a level or programming.

## REFERENCES

*Carbone, A., Hurst, J., Mitchell, I., and Gunstone D.. 2000. Principles for Designing Programming Exercises to Minimise Poor Learning Behaviours in Students. In Proceedings of the Australasian Conference on Computing Education (Melbourne, Australia) (ACSE '00). Association for Computing Machinery, New York, NY, USA, 26–33. https://doi.org/10.1145/359369.359374*

*Gamper, P., Schroeder, U.. 2023. Codescape - Development and Operation of a Serious Game for Teaching Programming in Introductory Courses. In: Haahr, M., Rojas-Salazar, A., Göbel, S. (eds) Serious Games. JCSG 2023. Lecture Notes in Computer Science, vol 14309. Springer, Cham. https://doi.org/10.1007/978-3-031-44751-8_26*

*Helminen, J., Ihantola, P., Karavirta, V., and Malmi, L.. 2012. How Do Students Solve Parsons Programming Problems? An Analysis of Interaction Traces. In Proceedings of the Ninth Annual International Conference on International Computing Education Research (Auckland, New Zealand) (ICER '12). Association for Computing Machinery, New York, NY, USA, 119–126. https://doi.org/10.1145/2361276.236130*

*Kasto, N., Whalley, J., Philpott, A., and Whalley D.. 2014. Solution Spaces. In Proceedings of the Sixteenth Australasian Computing Education Conference - Volume 148 (Auckland, New Zealand) (ACE '14). Australian Computer Society, Inc., AUS, 133–137.*

# Finding and Understanding an Impasse in Learning by Assembling with Epistemic Network Analysis

**Yusuke Hayashi**[1], **Ayano Ohsaki**[2], **Tatsuki Yamamoto**[3], **Tsukasa Hirashima**[1]
[1] Hiroshima University, [2] Shinshu University, [3] Meikai University
hayashi@lel.hiroshima-u.ac.jp

**ABSTRACT**: This study analyzes log data to confirm the occurrence and resolution of impasses in the learning process. Monsakun, a learning environment where learners create arithmetic word problems by combining given sentences, is examined in the context of university students. The investigation focuses on the occurrence of impasses through sequence analysis of log data. Epistemic Network Analysis (ENA) is used to model the connections in the data. The findings reveal that learners encounter impasses in level 3 exercises but gradually overcome them. The results demonstrate the potential effectiveness of ENA in investigating impasses and the use of learning by assembling to effectively evoke impasses in learning.

**Keywords**: Impasse in learning, Epistemic network analysis, learning by assembling

## 1 INTRODUCTION

Errors in learning can both lead to an impasse and promote learning (Tulis et al., 2016). Impasse-driven learning (VanLehn, 1988) is an effective way for learners to learn by identifying the causes of their impasse. Overcoming impasses is important in learning, and developing learning environments where learners can experience and overcome impasses on their own can be highly effective. However, the challenge is that many SA techniques are not easily interpretable, making it difficult for stakeholders to understand the data or know how to act on the results (Kleinman et al., 2022). In this study, we will discuss the analysis method to confirm the occurrence of impasse-driven learning in the learning process within a learning environment. To achieve impasse-driven learning, it is necessary to identify the occurrence of impasses and their causes, as well as detect the resolution of the impasses and the elimination of their causes. To address this, we utilize learning logs in a learning environment for problem-posing by assembling sentences, Monsakun (Hirashima et al., 2007), as well as Epistemic Network Analysis (ENA) (Shaffer et al., 2009).

## 2 METHODOLOGY

In Monsakun, learners create arithmetic word problems with given conditions by combining given single sentences. Figure 1 shows a screenshot of Monsakun. Learners are required to pose arithmetic word problems with provided sentences according to the requirements. Monsakun can automatically diagnose posed problems based on without natural language processing when learners ask to check their answer because it knows the conditions of arithmetic word problems and posed problems are limited in the combinations of provided sentences. This allows Monsakun to provide immediate feedback to learners. In Monsakun, learners must continue working on an exercise until they can correctly solve the problem. The target data in this study is the use of Monsakun by university students for a preliminary study prior to apply it to elementary school students as the main target. The large check count represents learners make many errors. Figure 2 shows the mean of check count of the answers. This result shows that the subjects come to an impasse from the first exercise of level 3 and then break it over at the end of the level. As shown in Table 1, each level has a different setting and has five exercises for each. From Level 3 the formula provided as the requirement is inconsistent with the required story type as shown in Fig. 1. This changes in the setting makes a muddle of their thinking. This study investigates what happens there with sequence analysis of log data.

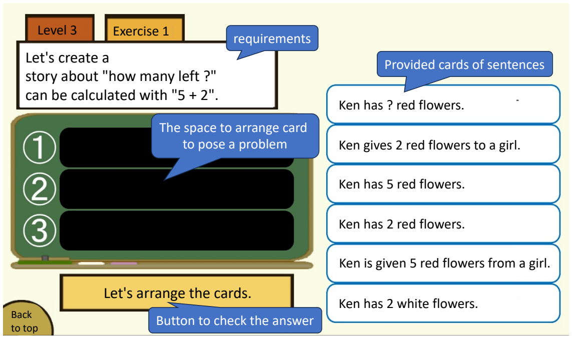



**Figure 1: Monsakun**

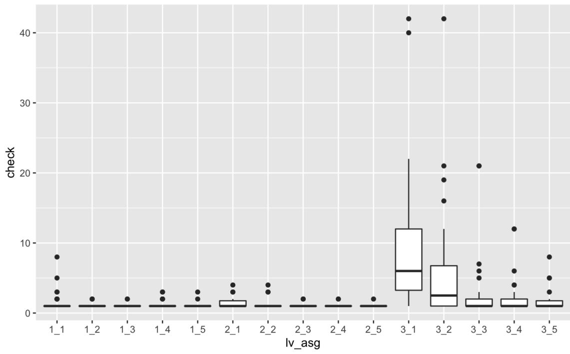



**Figure 2: Mean of check count of the answers**

**Table 1: Setting of levels in Monsakun**

| | *Level 1* | *Level 2* | *Level 3* |
|---|---|---|---|
| Mathematical formula | Story-based | Story-based | Solution-based |
| Consistency with words | Consistent | Inconsistent | Inconsistent |

ENA is a technique for modeling the structure of connections in data (Shaffer, 2017). In this study, we applied ENA to our data using the ENA Web Tool (version 1.7.0) (Marquart et al., 2018). We use the coding scheme shown in Table2. Learners' each response in Monsakun is combination of sentences. We label every response with a combination of code about the number order in the mathematical expression (F.*) and the sentence order for a type of story (S.*). For example, in Figure 1, when a learner arranges sentences "There are 8 boys in the park", "5 boys go home," and "There are ? (unknown number) boys in the park. (How many boys are in the park?)", this is labeled with "F.F_S.S".

**Table2: The coding scheme for log data**

| *Code* | *Description* | *Code* | *Description* |
|---|---|---|---|
| F.e | The response is empty | S.e | The response is empty |
| F.x | The response is *not* following the order of numbers in the mathematical expression. | S.x | The response is *not* following the sentence order for any types of stories. |
| F.F | The response is following the order of numbers in the mathematical expression. | S.s | The response is following the sentence order for different type of story requested. |
| C | Checked by learners and the answer is incorrect. | S.S | The response is following the sentence order for the same type of story requested. |
| E | The end of the exercise by the correct answer. | S.u | The response cannot be identified whether it is following the sentence order for any types of stories. |

## 3 RESULTS

Figure 3 shows the characteristic epistemic networks derived from the log data of levels 2-5, 3-1, and 3-5. All the means of the exercises at levels 1 and 2 are positioned almost in the same place close to the codes F.F_S.S and F.F_S.u on the right side. These codes mean learners try to follow both the mathematical expression and the sentence order required by story. On the other hand, means at the beginning of level 3, 3-1, 3-2, are scattered on the left top and ones at the end of level 3, 3-4 and 3-5, are scattered on the left bottom. This suggests that the subjects tend to follow the mathematical expression when solving problems at levels 1 and 2, but not as much at level 3, even though they attempt to construct the required story at all levels. Additionally, at level 3, the exercises are scattered from top to bottom. This indicates that at the beginning of level 3, there are many errors which gradually decrease towards the end of level 3. In level 3-1, there is a connection between "C" and "F.F_S.u". This suggests that the subject applies the same strategy used in levels 1 and 2, despite it not effective at level 3. This fixation can lead to an impasse.

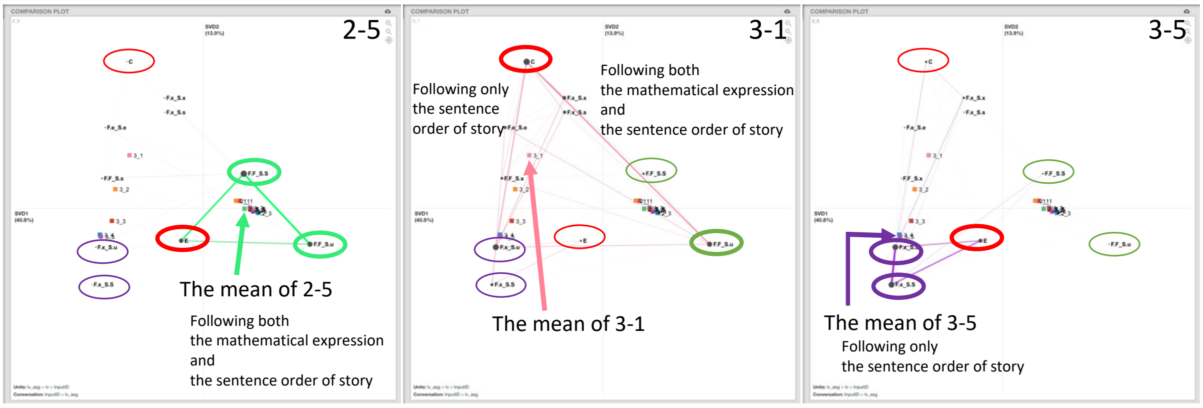



**Figure3: Epistemic networks of problem-posing exercises 2-5, 3-1 and 3-5**

## 4 CONCLUSION

This study examines the occurrence and resolution of impasses in Monsakun with ENA. ENA allows us to observe that from levels 3-1 to 3-5, the subjects encounter an impasse due to fixation on strategies from levels 1 and 2, but they can overcome it. Although this result is limited to a specific learning environment, it demonstrates the potential effectiveness of ENA in investigating impasses in the learning process and facilitating effective learning by triggering impasses through strategic assembly.


### ACKNOLWEDGMENT

This work was supported by JSPS KAKENHI Grant Number 23K11357.


## REFERENCES


Hirashima, T., Yokoyama, T., Okamoto, M., & Takeuchi, A. (2007). Learning by problem-posing as sentence-integration and experimental use. *AIED*, *2007*, 254–261.

Kleinman, E., Shergadwala, M., Teng, Z., Villareale, J., Bryant, A., Zhu, J., & El-Nasr, M. S. (2022). Analyzing Students' Problem-Solving Sequences. *J. of Learning Analytics*, *9*(2), 138–160.

Marquart, C. L., Hinojosa, C., Swiecki, Z., Eagan, B., & Shaffer, D. W. (2021). Epistemic network analysis (Version 1.7.0) [Software]. Available from http://app.epistemicnetwork.org

Shaffer, D. W., Hatfield, D., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E., Frank, K., Rupp, A. A., & Mislevy, R. (2009). Epistemic network analysis: A prototype for 21st-century assessment of learning. *International Journal of Learning and Media*, *1*(2).

Tulis, M., Steuer, G., & Dresel, M. (2016). Learning from errors: A model of individual processes. *Frontline Learning Research*, *4*(4), 12–26.

VanLehn, K. (1988). Toward a Theory of Impasse-Driven Learning. In H. Mandl & A. Lesgold (Eds.), *Learning Issues for Intelligent Tutoring Systems,* 19–41. Springer US.

# Analysing Learner Behaviour in an Ontology-Based E-learning System: A Graph Neural Network Approach

**Adam Wynn[1], Jingyun Wang[1], Zhongtian Sun[1], Atsushi Shimada[2]**
[1] Durham University, UK, [2] Kyushu University, Japan
adam.t.wynn@durham.ac.uk

**ABSTRACT**: Despite the prevalence of e-learning systems, there is a lack of support for learners to identify and compare new knowledge with existing cognitive structures. Therefore, an ontology-based visualization support system was previously introduced which offers two modes: cache-cache, where relations are initially hidden and the learners are encouraged to create those relations, and receptive, where learners can view expert-generated topic maps. In this study, we aim to analyse learner behaviour by representing user behaviour as graphs and utilising a heterogeneous graph convolutional network. Two graphs are constructed for each student to capture behaviour before and after system use. Results indicate significant differences in mean embeddings between learners in receptive and cache-cache modes. Further analysis, considering pre-test performance, shows no significant differences in the receptive and cache-cache groups but highlights a considerably smaller mean for high prior performers in the cache-cache group.

**Keywords**: Meaningful learning; discovery learning; ontology; topic-map; graph neural network

## INTRODUCTION

Learning strategies that support the organisation of knowledge within a hierarchical cognitive framework is highlighted by the hierarchical nature of knowledge significantly increases the learning performance of learners (Tsien, 2007). Understanding and making connections between new material and pertinent concepts are key components of meaningful learning, which is described as the substantive integration of new concepts into pre-existing cognitive frameworks (Ausubel, 1963; Ausubel et al., 1978). Despite having started to replace traditional textbooks with digital ones, existing e-learning systems lack support for learners to identify and compare new knowledge with existing cognitive structures (Wang et al, 2019a; Wang et al, 2020).

In Wang et al (2020) the authors introduce an ontology-based visualization support system designed for e-book learners, fostering both meaningful receptive learning and meaningful discovery learning. The system has two learning modes: (1) cache-cache mode where to begin with, all information regarding relations is hidden and the learners are encouraged to discover them, and (2) reception comparison mode where learners can see complete versions of expert generated topic maps. In this paper, we aim to analyse the behavioural differences among students when using cache-cache mode and receptive modes by using ontology data and log data obtained before and after the learner used the system in an existing computer science course. Analysing the learner data can reveal how the different learning modes in the system enable more effective learning of new concepts and how students explore the relationships between concepts during learning.

## METHODOLOGY

This work proposed representing the user's behaviour as a graph and input into a heterogenous graph convolutional network (Wang et al, 2019b) using PyG (Fey and Lenssen, 2019) to take advantage of the multiple node and edge types. Two graphs are constructed for each student: one which represents their learning behaviour before using the system, and one for their behaviour after using the system. As shown in figure 1, each graph has page nodes which represent each page in the book, knowledge point (KP) nodes defined as "a minimum learning item which can independently describe the information constituting one given piece of knowledge in the content of a specific course" (Wang et al, 2020), and their upper concept nodes. There are edges linking each page node if the student moved to that page and stayed between 4 and 1200 seconds where the edge weight denotes the number of seconds spent on the previous page. Each page is linked to a KP if that KP is found on the page, and each KP is linked to an upper concept if the KP belongs to that upper concept.

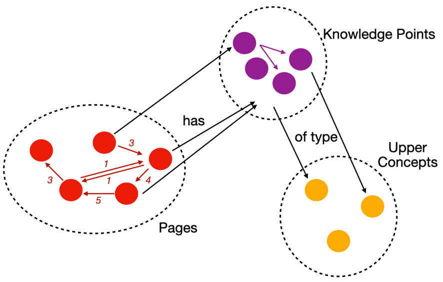



**Figure 1: Representation of the graph constructed for each student**

After the graphs are constructed, the model is trained on the task of link prediction between page nodes using Adam Optimizer (Kingma et al, 2014) and binary cross entropy loss. After training, three types of embeddings are derived from the last hidden layer which act as mathematical representations of each type of node in the graph. These embeddings represent how students interact with the e-books and capture nuances in their learning behaviour. Each student has two sets of page, KP, and upper concept embeddings representing their behaviour before and after using the system.

## RESULTS

To analyse the effectiveness of the system, embeddings for each student before and after using the system, as well as the embeddings for the cache-cache and receptive modes are compared. In this study, we focus on comparing the average page embeddings of each student, as they best capture the behaviour of students compared to other node types. Embeddings are passed through a multilayer perceptron to be reduced to one dimension. Figure 2 shows the distribution of these embeddings.

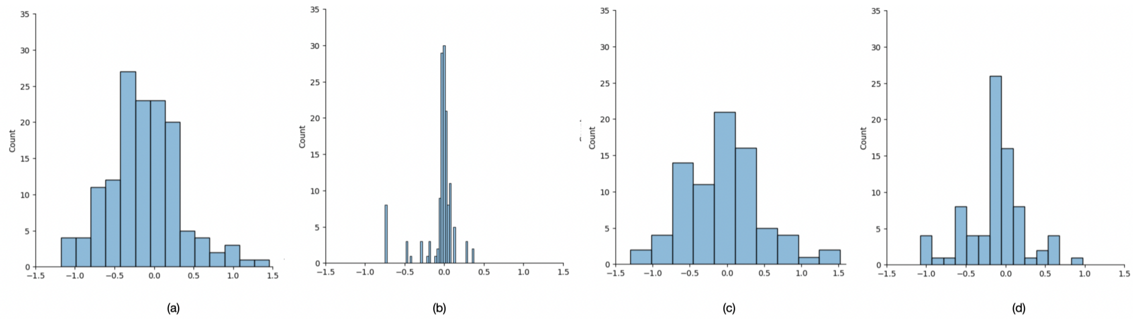



**Figure 2 – Distribution of page embeddings for (a) Receptive Mode Before, (b) Receptive Mode After, (c) Cache-Cache Mode Before, (d) Cache-Cache Mode After**

In Figure 3, it can be seen that the embeddings which represent learner behaviour after using the system are more clustered around the centre compared to the embeddings before using the system which are more spread out. More specifically, the "Receptive After" nodes are more clustered than the "Cache-Cache After" nodes. This could suggest that learner behaviour is more consistent with each

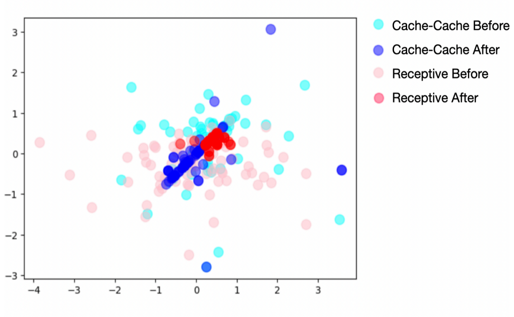



**Figure 3 – Plot of Page Embeddings**

other after using the system, especially for receptive mode indicating the system potentially helped students converge their behaviour on effective learning strategies.

Furthermore, a Mann-Whitney U test was performed to evaluate whether the behaviour differed by modes. The results indicated the mean page embeddings were significantly different ($z=-2.582$, $p= 0.01<0.05$) between learners in receptive mode and in cache-cache mode. Page embeddings were also analysed grouped by previous knowledge determined by pre-test. For the receptive group, the group with high performance (Mean = 0.057, S.D.= 0.212) and the group with low performance (Mean = 0.061, S.D. = 0.182) do not show significant differences. ($F(1,78)=0.010$, $p=0.919$), and for the cache-cache group, the group with high performance (Mean = 0.038, S.D. = 0.348) and low performance (Mean = 0.112, S.D. = 0.299) also do not show significant differences ($F(1,78)=1.042$, $p=0.310$), but the mean for the group with high performance is much smaller than the group with low performance.

## CONCLUSION AND FUTURE WORK

In summary, this study demonstrates how log data and ontology data can be used as an input to a graph neural network to produce graph embeddings which can represent learner behaviour. Currently, page embeddings are analysed by considering the mean embeddings for all pages for each student and by reducing them to one dimension, therefore, some information in the data may be lost. In the future, we will analyse page embeddings together with KP and upper concept information in more detail by considering individual embeddings for each page.

## ACKNOWLEDGEMENTS

This work was supported by JST CREST Grant Number JPMJCR22D1 and JSPS KAKENHI Grant Number JP22H00551, Japan.

## REFERENCES

Ausubel, D. (1963). The psychology of meaningful verbal learning. Grune & Stratton.

Ausubel, D., Novak, J., & Hanesian, H. (1978). Educational psychology: A cognitive view. Holt, Rinehart & Winston.

Fey, M., & Lenssen, J.E. (2019). Fast Graph Representation Learning with PyTorch Geometric. ArXiv.

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. arXiv.

Tsien, J. Z. (2007 July). The memory. Scientific American, pp. 52–59.

Wang, J., Shimada, A., & Okubo, F. (2019). E-book learner behaviors difference under two meaningful learning support environments. ICCE 2019 - 27th International Conference on Computers in Education, Proceedings (pp. 342-347).

Wang, X., Ji, H., Shi, C., Wang, B., Ye, Y., Cui, P., & Yu, P. S. (2019). Heterogeneous graph attention network. In The world wide web conference (pp. 2022-2032)

Wang, J., Shimada, A., Oi, M., Ogata, H. & Tabata, Y. (2020). Development and evaluation of a visualization system to support meaningful ebook learning. Interactive Learning Environments

# Multiple-Choice Question Generation Using Large Language Models that Also Controls Discrimination

**Yo Ehara**
Tokyo Gakugei University
ehara@u-gakugei.ac.jp

**ABSTRACT**: In question generation for education, controlling the difficulty of the questions according to the learner is important. Many studies have been conducted to address this issue. However, discrimination is also important in estimating learner ability using as few questions as possible. A question with high discrimination would accurately identify learners with a high ability to answer that question. Discrimination has rarely been addressed in question generation using large language models. In this study, we propose a method for generating questions with controlled discriminative power by prompting large language models.

**Keywords**: Large language models, Discrimination, Item Response Theory

## 1 INTRODUCTION

Generative artificial intelligence (AI), which can generate questions according to specific instructions, has attracted considerable attention. Considering that questions can be generated by specifying the difficulty to the generative AI, a method for specifying the difficulty scale independent of the field would be useful. Field-independent measures include difficulty and discrimination measures based on item response theory (IRT), which necessitates that the human instructor be familiar with statistical measures such as IRT [Baker, 2004]. Discrimination measures how well the question discriminates high-ability learners from low-ability learners based on correctly answered questions.

Previous studies that generated educational questions for language learning are summarized in [Cui et al., 2023], . However, these studies, including [Cui et al., 2023] do not focus on the IRT discrimination parameter [Baker, 2004]. Farr [2024] recently studied methods of generating English vocabulary questions for learners with English as a second language, although discrimination was not mentioned.

Therefore, we propose a method to control discrimination in prompts to large language models. . Specifically, we propose a method that uses the distribution of the number of correct respondents as a measure that is easier for both humans and AI to understand: "In a group of 100 examinees, the following question is expected to have a mean of a and standard deviation of b in the predicted distribution of the number of correct respondents. Please generate a completely new question with a predicted distribution of mean c for the number of correct answers." By providing such instructions in prompts, teachers can easily specify the difficulty and discrimination of the question, even if they are unfamiliar with statistical scales and do not understand the concept of IRT discrimination. Through experiments, we show that specifying the standard deviation of the examinees in prompts has an effect similar to that of specifying the value of the IRT discrimination parameter in prompts.

## 2 GENERATING MULTIPLE-CHOICE QUESTIONS

We conducted an experiment to generate a new English test for these questions using GPT-4 (ChatGPT May 24 version), a well-known generative AI. In most cases, GPT-4 successfully generated meaningful and answerable vocabulary questions in several attempts. Specifically, the following instructions were used to specify the types of questions the user wanted to generate: The actual prompts were specified in Japanese.

Actual Prompt used: The following English vocabulary test question is predicted to have a mean distribution of 69 correct answers with a standard deviation of 4.16 for a given group of 100 English language learners. For this population of test takers, generate English vocabulary test questions with a similar distribution of predicted number of correct answers and smaller standard deviation, using completely different words. ====="The area was ______ in timber and coal." Choose one of the following four words from the underlined words: expensive, cheap, poor, or not well off =====

Hereafter, we refer to the question between "=====" as the original question. Japanese was used for prompting because the dataset of [Ehara,2022] was created by Lancers, a Japanese crowdsourcing service, and the majority of respondents were assumed to be English learners whose native language was Japanese. From the viewpoint of respondents' privacy, we did not ask respondents to input their native language, nor did we explicitly allow only native Japanese speakers to answer the questionnaire, which can only be inferred or implied. Therefore, generating English word test questions with Japanese instructions in the actual prompt, implies that the test questions are generated for learners of English with Japanese as the native language. No explicit instructions were provided in the prompts for generating test questions for native English speakers of Japanese.

Instructed to generate a question so that the number of correct answers is 95 out of 100. He turned off the lights and went to ______. a) bed, b) kitchen, c) car, d) park

Instructed to generate a question so that the number of correct answers is 29 out of 100. The researcher analyzed the _____ between the two variables. a) apex, b) correlation, c) paradigm, d) zenith

Instructed to generate a question so that the standard deviation is larger than that of the original question: He is very ______. a) kind, b) friendly, c) evil, d) generous

Instructed to generate a question so that the standard deviation is smaller than that of the original question: The birds ______ south for the winter. a) fly b) drive c) walk d) run

In the case of the standard deviation of the distribution of the number of correct answers, no specific numerical target was specified, but simply "large" or "small" because it is difficult even for humans to predict the standard deviation of the created questions. When the standard deviation of the distribution of the number of correct answers was made large, questions in which more than one option is considered correct were generated. However, when the standard deviation of the distribution of the number of correct answers was made smaller , questions with more obvious correct answers were generated.

Next, we investigated another research question: if the same instructions are given in technical terms such as "difficulty" or "discrimination" in IRT, would the GPT-4 correctly understand what kind of questions to generate? Therefore, we generated questions using these terms.

Instructed to generate a question so that the IRT difficulty parameter would change from -1.2 of the original question to 0.89. Despite the challenging conditions, the team remained ______. a) resilient b) permeable c) solvable d) inflammable

Instructed to generate a question so that the IRT discrimination parameter would be smaller than 0.738 of the original question: The sun ______ in the east. a) rises, b) falls, c) sinks, d) dives

Instructed to generate a question so that the IRT discrimination parameter would be larger than 0.738 of the original question: He has a ______ to exaggerate things. a) tendency b) progression c) direction d) development

When instructed to generate a question with greater discrimination, the difficulty level was increased by using low-frequency words. When instructed to generate a question with less discrimination, the generated question was more ambiguous, particularly when low-frequency words were not used. Surprisingly, GPT-4 also appears to have an adequate understanding of IRT concepts such as difficulty and discriminability and is qualitatively capable of generating questions in line with the instructions.

# 3 CONCLUSIONS

In this paper, we proposed a method for generating multiple-choice questions (MCQs). First, we presented a sample question with the mean and standard deviation of the number of learners who correctly answered the question. Then, GPT-4 was prompted to generate a new question with a different mean or standard deviation, observing qualitatively that the intended questions were generated. Subsequently, we checked whether the same technique would work using technical terms such as difficulty and discrimination instead of mean and standard deviation. Surprisingly, GPT-4 seemed to understand these technical terms and generated questions with the intended characteristics. Hence, both methods can be used to control discrimination when generating MCQs. Although we used MCQs in English vocabulary tests for second-language learners only, the proposed methodology can be easily applied to MCQs in other fields because the user only needs to change the questions in the prompts. Future work will include a comprehensive analysis of these questions.

**Acknowledgements:** This work was supported by JST ACT-X JPMJAX2006, and JSPS KAKENHI 22K12287.

# REFERENCES

Ehara, Y. No meaning left unlearned: Predicting learners’ knowledge of atypical meanings of words from vocabulary tests for their typical meanings. Proc. of EDM (short paper), 2022.

Baker, F.B., Item Response Theory : Parameter Estimation Techniques, Second Ed., CRC Press, 2004.

Cui, P. and Sachan, M. Adaptive and Personalized Exercise Generation for Online Language Learning. In Proc. of ACL, pp. 10184–10198, Toronto, Canada. 2023.

Farr, C. Unmasking ChatGPT: The Challenges of Using Artificial Intelligence for Learning Vocabulary in English as an Additional Language. Master Thesis. Univ. of Victoria. 2024.

# Decoding Ambition Metrics: Unveiling the How, Why and Neuroscience of Automated Feedback

**Etain Kiely, Ikechukwu Ogbuchi, Cormac Quigley, Donal McGinty, Konrad Mulrennan, John Donovan**
Atlantic Technological University, Ireland.
Etain.kiely@atu.ie

**ABSTRACT**: This demonstration unveils an innovative, scalable, ground-up system for implementing automated personalized feedback. This approach is informed by neuroscience and offers educators flexibility to provide constructive, encouraging feedback aligned with learners' success and engagement threshold metrics derived from the VLE environment. The analysis explored navigation behavior and interactions to gain insights into self-regulated learning strategies (Hadwin & Järvelä, 2020). Metacognitive prompts were used as learning scaffolds to help students improve their self-regulation and encourage active learning, rehearsal, appraisal, and elaboration (Nguyen et al., 2020). Learners played a pivotal role in shaping the project's development, providing invaluable insights into optimal feedback timing, mechanisms, and language. The learning analytics research cycle was employed to iteratively evolve and streamline the system over eight years using MS Power Automate, Python, and APIs. The output is a personalized email sent to over 400 first-year math students. offering personalized learning paths, timely interventions, optimized instructional strategies, enhanced motivation, and targeted feedback.

**Demonstration Link:** https://www.youtube.com/watch?v=7j6Ey19T5_Q

**Keywords**: Personalized Feedback, Machine Learning and NLP, Retention Strategies

## 1 KEY FINDINGS

The impact of this intervention was monitored through changes in learner log interactions and qualitative feedback. Application of machine learning models, particularly k-means clustering, effectively identified students at risk as early as week 5. This research has paved the way for the next phase of research, which aims to scale this innovative approach across the entire university. This will involve in-depth analysis of learners' processes, examining sequential actions using advanced learner modeling techniques such as Hidden Markov Models (HMMs). By exploring ambition and success metrics, educators can not only observe and monitor learners' behaviors but also establish clear success thresholds and provide timely, inclusive, and adaptive support to optimize learning.

## REFERENCES

Hadwin, A. F., & Järvelä, S. (2020). Learning analytics to reveal links between learning design and self-regulated learning. Learning and Individual Differences, 83, 101890.

Nguyen, T. B., Koper, R., & Olivier, B. (2020). Discovering the Effects of Metacognitive Prompts on the Sequential Structure of SRL-Processes Using Process Mining Techniques. Journal of Learning Analytics, 7(1), 29-52.

# ART-Math: A Story-based Creative Math Learning Platform

**Yukyeong Song**
University of Florida
y.song1@ufl.edu

**Wanli Xing**
University of Florida
wanli.xing@coe.ufl.edu

**Chenglu Li**
University of Utah
chenglu.li@utah.edu

**ABSTRACT**: Traditional math education focuses on the "right" answer instead of exploring processes and creative approaches. To change this, we developed an interactive learning platform called ART-Math (Ask, Representation, and Transformation), which provides a story-based creative learning experience for elementary school students. We field-tested with an authentic use case with twelve 5th-grade students in an elementary school in the Southeastern part of the United States. We received real-time feedback from the students and teachers during the usability testing and conducted a pre, post knowledge test and an engagement survey. The paired-sample T-test results show a significant improvement in their mathematics knowledge (multiple-choice: $t(11) = 2.449$, $p < 0.05$, open-ended: $t(11) = 3.954$, $p < 0.05$). The engagement survey shows that students enjoyed the ART-Math class ($M = 4.75$, $SD = 0.45$), and found the platform easy to use ($M = 4.42$, $SD = 0.67$). We found a promising potential for ART-Math to introduce learning analytics research by using students' interaction data with the platform and the generative AI as well as the design of the teacher dashboard. We demonstrate ART-Math in this link.

https://drive.google.com/file/d/1YOuUVJEHtpWth4RvBbCX6PdUAqQzdZU-/view?usp=drive_link

**Keywords**: Math learning platform, interactive learning technology, learning analytics

# An artificial intelligence teachable agent to support learning-by-teaching of secondary school students

**Yukyeong Song**
University of Florida
y.song1@ufl.edu

**Chenglu Li**
University of Utah
chenglu.li@utah.edu

**Wanli Xing**
University of Florida
wanli.xing@coe.ufl.edu

**ABSTRACT**: Scholars supporting Social-cultural constructivism have emphasized the importance of learning through social interactions. Stemming from this learning theory, the theory of learning by teaching has emerged. Learning by teaching provides learning opportunities by allowing students to play the role of teachers, mostly realized in such forms as peer learning. Recently, a notable advancement in AI technologies, such as large language models, has made it possible to replace human peer roles with an AI teachable agent. Theoretically grounded on the learning-by-teaching framework with components of gamification, we developed the ALTER-Math (AI-augmented Learning by Teaching to Enhance and Renovate Math Learning) platform. ALTER-Math is designed for secondary school learners to learn mathematics topics by teaching the AI teachable agent. The agent demonstrates a not-so-knowledgeable peer student and visualizes the knowledge growth through points and dashboards. Following the design and development research method, we have developed the initial working prototype of ALTER-Math and completed the first round of the user study with eight teachers. The strengths and points of improvement have been identified through the user studies and interview responses, and our team is expected to develop the second prototype of the ALTER-Math. The demo of ALTER-Math is found here.

**Keywords**: large language model, AI teachable agent, learning by teaching, mathematics learning

HTTPS://WWW.YOUTUBE.COM/WATCH?V=710GU2KILWK

# Enhancing Self-Regulated Learning Through Personalized Analytics

**Jonathan Hahn, Ursina Raemy, Kinga Sipos, Sarah Schnyder, Stefan Troche, Natalie Borter**
University of Bern
jonathan.hahn@unibe.ch, ursina.raemy@unibe.ch, kinga.sipos@unibe.ch, sarah.schnyder@unibe.ch, stefan.troche@unibe.ch, natalie.borter@unibe.ch

**ABSTRACT:** Our demonstration presents a learning analytics tool designed to enhance self-regulated learning among university students through formative assessments and feedback mechanisms. The tool features a user-centric interface, leveraging personalized data to provide real-time, tailored feedback and actionable recommendations, especially during students' self-learning phases. The tool has been field-tested at the Institutes of Psychology, Sports Science, and Mathematics of the University of Bern. We have collected feedback from over 1000 students and teachers through interviews and online questionnaires. This feedback has led to continuous improvements. Based on the insights gained, we have developed additional formative assessments for each content area and a dashboard that displays individual performance and time investment. Furthermore, the feedback mechanisms for students address not only content-related aspects, such as response correctness and explanations but also offer personalized guidance on improving self-regulated learning, like advice for solving additional exercises or tips for more efficient and successful learning strategies. The demonstration showcases both the early-stage prototype, which consisted solely of formative assessments, and its evolution into a mature, user-centric product, emphasizing its effective enhancement of the educational experience.

**Keywords**: learning analytics tool, self-regulation, formative assessments, feedback

## DEMO VIDEO

https://youtu.be/E74XORbFZGY

## ACKNOWLEDGEMENTS

This work was supported by a BeLEARN grant.

# Demo of TMR: a System that Encourages Reflection Using Text Mining

**Yuichi Nukiyama**[1]**, Toshihiro Kita**[2]
Kumamoto University, Japan
nukiyama.yuichi@gmail.com[1], kita@rcis.kumamoto-u.ac.jp[2]

**ABSTRACT**: Since the onset of the COVID-19 pandemic, many corporate training programs have transitioned to an online format. This shift has brought about the advantage of enabling learning without constraints of location or time. However, it has also raised concerns regarding disadvantages such as the propensity for passive engagement in training, limited interaction among learners, and challenges in facilitating discussions. Consequently, this paper aims to address the difficulties associated with online training, particularly in supporting discussions. This paper focuses on facilitating reflection within synchronous online training discussions using Zoom. Zoom has a feature for transcribing audio discussions into text format. By subjecting this text to various text mining techniques for visualization, there is potential to support learners' reflection. To achieve this, a system named Text Mining for Reflection (hereinafter referred to as TMR) was designed and developed. TMR is integrated as an application within Zoom Apps, accessible and operable seamlessly from the Zoom desktop client during online training sessions. The text mining functionalities of TMR include: 1. Discussion summarization using GPT-3.5, 2. Visualization of word relationships through co-occurrence network diagrams, 3. Word cloud representation utilizing tf-idf, 4. Word cloud depiction of frequently used terms, and 5. Scatter plot visualization with Word2vec. In the assessment of TMR, both heuristic analysis and system testing have been conducted. The heuristic analysis involved evaluative feedback from analysts regarding the system's usability. Concurrently, analysts raised concerns that users might find it challenging to comprehend the text mining methods. Consequently, a tooltip feature has been added to the text mining method selection screen to address this issue. The system testing was conducted in the context of a discussion focused on improving lesson plans. Through this testing, confirmation was obtained that TMR delivers the functionalities it was designed for. Additionally, positive comments were received from test participants, highlighting the introspective benefits of visualization through text mining and the ability to understand the roles each participant played in contributing to the discussion.

**Keywords**: Reflection, Discussion, Text Mining, Online Classroom, Zoom Apps, Learning-support System

**Video:**
https://drive.google.com/file/d/10XBxGcHxU3mvW_y2RP0nRCiOKGn7LEMF/view?usp=sharing

# UniAnalytics: A Learning Analytics Tool to Support Teaching and Learning with Jupyter Notebooks

**Zhenyu Cai, Richard Davis, Raphaël Mariétan, Pierre Dillenbourg**
École Polytechnique Fédérale de Lausanne
{zhenyu.cai, richard.davis, raphael.marietan, pierre.dillenbourg}@epfl.ch

**ABSTRACT**: UniAnalytics is a pair of extensions tailored for teaching and learning in the JupyterLab environment. By providing a seamless interface for instructors to observe group dynamics during lectures and exercise sessions, Unianalytics enhances the interactive capabilities of Jupyter Notebooks. This user-friendly tool operates as a simple package, easily installable in any JupyterLab environment. In our demo walkthrough, we illustrate Unianalytics in action using the same notebook for teachers and students. From setup to data collection using the extensions, we showcase the teacher dashboard that can be used to monitor both synchronous and asynchronous learning activities.

**Keywords**: Learning Analytics, Jupyter Notebooks, Classroom Orchestration, Higher Education

## 1 OVERVIEW

UniAnalytics comprises of two components, the *Telemetry* on the student side, and the *Dashboard* on the teacher side. *Telemetry* anonymously logs the student's interaction with the notebook at a finer granularity. For example, every execution or click on both the code and Markdown cells is captured while the content and timestamp are recorded as well. This enables the data visualization on the *Dashboard*, including the Table-of-Content (TOC) dashboard on the left sidebar, as well as the notebook-level and cell-level dashboards on the right sidebar. The teacher can use the TOC dashboard to track "where students are" in real time according to the number of students working on each part of the notebook. This feature can be enhanced by the cell-level dashboard, on which the teacher can delve into the mistakes that students make by checking their inputs and/or outputs. By doing this, the teacher can make data-driven decisions and adapt the lesson plan to the classroom dynamics, i.e., performing classroom orchestration. Moreover, the notebook-level dashboard provides an aggregate view of how the whole class performs on this notebook, including the number of clicks and executions and the time students spend on each cell. This can be used to estimate the class engagement as well as the difficulty level of the learning materials, and thus to support notebook revision to improve the learning outcomes. In addition to the above synchronous settings, the teacher can also revisit the historical data visualized by the *Dashboard* or export the raw data for further research.

## 2 DESIGN AND EVALUATION

The design of UniAnalytics followed a design-based research method, incorporating human-centered learning analytics approaches. By interviewing end users with a preliminary mockup, we re-designed the dashboard, and implemented it into the current version. The tool has been tested in several university STEM courses, in one of which the dashboard indicated that there were at most 43 students simultaneously working on the notebook. Continuing with this co-designing process in authentic settings, we aim to bring more insights from stakeholders and develop features to address their needs of learning analytics in higher education.

## 3 DEMO VIDEO

For demonstrating how UniAnalytics works, we provide a video at: https://youtu.be/y5GKMu7PUw8.

# Accelerating Platform-Based Learning Science Experimental Research

**Ilya Musabirov**
University of Toronto
imusabirov@cs.toronto.edu

**ABSTRACT**: In my thesis project, I design algorithms and tools, integrated with educational data repositories and data analysis instruments, focusing on accelerating scientific discovery with educational experimentation platforms while supporting instructors in helping students. The proliferation of blended education during the last few years dramatically increased both in higher education and in K-12, with educational platforms reaching hundreds of thousands of students, and expanding opportunities for behavioral and instructional interventions. However, some specific characteristics of educational settings restrict the applicability of both traditional experimental research and industrial AB-comparisons approaches to the problem. Two key directions of the project are exploring novel ways to aggregate heterogenous evidence and supporting instructors in choosing evidence-based interventions while mitigating risks driven by the heterogeneity of treatment effects.

**Keywords**: knowledge accumulation, experimental research, metaanalysis, contextual multiarmed bandits

## 1 PROBLEM BACKGROUND AND STUDY CONTEXT

The proliferation of blended education has dramatically increased during the last few years both in higher education and in K-12 with platforms such as CMU Open Learning Initiative or Carnegie Learning MATHia, as well as traditional LMS systems and online courses, reaching hundreds of thousands of students.

Acknowledging emerging affordances for large-scale learning and behavioral science research, as well as the need for knowledge accumulation for continuous improvement of educational practice, policy setters announced a range of "challenges" and grant opportunities, aimed to support the creation of educational research infrastructures at scale. Some examples of such initiatives in North America are the XPRIZE Digital Learning Challenge (DLC), the National Science Foundation Cyberinfrastructure and Research Infrastructure grant calls, the Institute for Education Sciences SEERNet, and the Learning Agency's Tools Competition.

One of the foci of these initiatives is the creation or scaling up of digital educational experimental platforms (DEEP), delivering on the promise of "super experiments" (Stamper et al., 2012), outlined more than ten years ago.

There are multiple ongoing attempts to develop DEEP, for example (Fancsali et al., 2022; Heffernan & Heffernan, 2014; Motz et al., 2018; Reza et al., 2021), as well as a larger body of work on AB-testing in technology industry settings (Deng et al., 2017) with attempts to apply these approaches to education (Sankaranarayanan et al., 2023).

While on the basic level, DEEP are not unlike their AB-testing counterparts in industry, their direct promise is to build more rigorous learning science evidence. Moreover, there are specific characteristics of educational settings that are the reality experimental platforms need to deal with. Diverse audiences, sample sizes limited by enrollment numbers and ranging 20-1000 students per section, restrict the applicability of both traditional experimental research and industrial AB-comparisons approaches to the problem (Musabirov, 2022). Some of these issues are already being addressed in educational experimentation platforms (Fancsali et al., 2022; Heffernan & Heffernan, 2014; Motz et al., 2018; Reza et al., 2021). In the current work, however, I focus on two underexplored aspects of the current educational landscape which set new requirements for DEEP: accounting for the heterogeneity of treatment effects of interventions and balancing research and continuous improvement to cater to interests of different stakeholders, allowing them to share the focus with researchers.

There is growing recent attention in learning and behavioral science communities is drawn by the apparent need to explore and understand the heterogeneity of treatment effects of interventions (Bryan et al., 2021; Kizilcec et al., 2020), which is possible only at scale. In other words, we need to assume that the effects of many interventions will vary in different contexts and need to account for that. For human-computer interaction and educational technology communities, this, in turn, poses additional challenges to learning to act based on the observed heterogeneity when translating knowledge to field interventions (Chen et al., 2022).

Moreover, the tight integration of DEEP in courseware systems calls for reevaluating how we account for the interests of different stakeholders, for example, learning designers, instructors, and students. One of the lenses is the trade-off between acquiring scientific evidence and more rapid deployment of results in the classroom. Machine learning-enabled adaptive experiments are one of the promising ways to navigate this trade-off (Reza et al., 2021), allowing to combine learning from interaction with experiment participants with ongoing continuous improvement by reallocating future interactions to better options.

The context of my thesis study is a research project “Frameworks for Intelligent Adaptive Experimentation: Enhancing and Tailoring Digital Education,” dedicated to designing and building this future-generation adaptive experimental infrastructure, which employs machine learning algorithms for adaptive experimentation.

## 2 RESEARCH GOALS AND METHODOLOGICAL APPROACH

While there is still much work to be done in understanding and developing ways to design these platforms, my thesis starts from a speculative proposition. Imagine, that we are already at the point where hundreds of thousands of experiments are performed with the help of DEEP in our blended classrooms. What are some ways they can change the practice of learning science research and learning design, taking into account the interests of research, policy, and learning design/teaching stakeholders?

My choice heuristic is on the points where we need to balance the needs of researchers, instructors, and students in exploring trade-offs between accumulating scientific knowledge and supporting the rapid improvement of learning.

In particular, I look at the cases where the design of DEEP can:

- Support researchers and instructors in intervention design and data analysis of the evidence gathered by DEEP, accounting for heterogeneity;
- Support instructors in choosing evidence-based interventions to adopt in their classrooms and ease adoption when there is some evidence and mitigate risks driven by heterogeneity of treatment effects.

# 3 STUDY OUTLINE

This study outline represents the partially ongoing work. I focus on the main ideas and state of each prototype, including the current progress and open questions. For each of the prototypes I follow the traditional iterative design cycle, including research involving stakeholders, series of prototypes of different fidelity and evaluation.

## 3.1 Methods and tools supporting intervention design and data analysis of the evidence gathered by DEEP, accounting for heterogeneity

Such components of DEEP as unified educational data repositories and integrated tools for data analysis will be used to support best intervention design and data analysis practices by providing code- and visual workflows. These workflows will support model-based inference for planning and analyzing interventions and aim to improve data visualization and communication of experimental results.

Another critical opportunity for DEEP in this direction is supporting intelligent data aggregation from multiple studies to account for effect heterogeneity. The current state-of-the-art approaches to evidence gathering and systematization are backward-looking – What has been observed? Answering this question requires a manual, high-cost meta-analytical effort limited by the need to account for diverse statistical designs and approaches (Kale, 2023). There are new opportunities to be forward-looking in the sense of using data from experiments to rapidly impact the future experiments' design. This can lead to setting better standards of evidence: having hundreds of courses in different institutions in parallel, having them repeatedly for many terms gives good source data for disentangling many sources of variation not distinguishable in traditional meta-analytic models, as well as accounts for heterogeneity, which is the key to efficient behavioral science interventions in education.

Methodologically, this work will be based on hierarchical probabilistic Bayesian models, because they are well suited to pooling data from many small and diverse samples. Some open questions that will require research contributions in applied statistics and ML engineering are:

1. Parallel fitting and evaluation of competing models, with different pooling/and sets of covariates with champion-contestant evaluation,
2. Ensuring always-valid inference and other ways of supporting decision-making without violating false positive rate guarantees,
3. Finding ways to balance knowledge capitalization goals, supported by multi-armed bandits, and ensuring valid inferences from aggregating multiple sources of bandit-collected data.

By bringing these capabilities to trusted research platforms, I expect to alleviate some of these potential problems, but I anticipate the need to focus on more automated analysis workflows, supporting continuous improvement.

### 3.2 Developing portable context-aware intervention formats palettes

Based on the evidence gathered using DEEP, how can we ease the classroom adoption of portable interventions, while mitigating risks driven by heterogeneity of treatment effects for the particular classrooms and decreasing the burden of manual instructor's decision making? As one of the potential answers, I explore the format of a portable executable adaptive intervention based on bridging behavioral science-informed content with a contextual bandit design.
The key idea is to explore the match between the contextual effects of interventions (Kizilcec et al., 2020) and opportunities provided by contextual multiarmed bandits (cMABs) (Bouneffouf et al., 2020).

cMABs are online machine learning algorithms that take into account contextual variables, representing e.g. the properties of student or task to guide ongoing continuous improvement by reallocating future interactions to better options. For example, they can learn to allocate particular intervention condition to students who made particular errors on the previous task if it works better for them and apply accumulated evidence to continuously improve the assignments.

This match can help the designers and instructors to adapt evidence-based interventions as safe bets even in cases when real behavior of students in the course deviates from what we learned so far due to unexpected heterogeneity: while we start based on already accumulated evidence, cMABs learn online and can correct if the reaction of students differs from those discovered in previous studies. This will provide learning designers and instructors with opportunities to quickly adopt relevant learning or behavioral interventions based on current accumulated knowledge and execute them in a safer adaptive way.

One ongoing focus is on investigating how the contextual bandit algorithm can incorporate alternative priors. I will investigate how these can best reflect the current knowledge from contextual variables, representing student and course characteristics from experiments across multiple courses, in initializing algorithms and analysis for new deployments and replications.

## 4 CURRENT AND EXPECTED CONTRIBUTIONS

- Field experiments using adaptive and non-adaptive factorial designs for instructional (explanation vs self-explanation) and motivational (encouraging students to solve optional problems) goals
- Prototype architecture for integrating courseware systems with adaptive experimentation, developed and tested within five rapid replications during the XPRIZE DLC
- Simulation-based guidelines to estimate potential achievable rewards for the range of most common intervention and interaction effect sizes to outline when applying the standard bandit algorithms might be practically valuable in educational settings

- Patterns of adaptive context-based behavioral interventions, designed and evaluated during the XPRIZE DLC, confirming the feasibility of developing motivational interventions transferable between courses.

This work enables the incorporation of context-aware knowledge of what works for whom based on the current evidence, as well as the opportunity for personalization, adaptation, and accounting for error. I will explore how to identify when the intervention works differently in the context of a particular course. I will use these insights to explore how to set parameters of bandit algorithms, so they automatically recalibrate what probability we expose which students to which arms. To close the loop with behavioral scientists, I will investigate what data to present and how to communicate analyses via a live meta-analysis model, with a focus on directing researchers' and instructors' attention to which aspects of the intervention are best suited for redesign and improvement.

Developed tools and prototypes will help learning scientists and educational researchers advance our understanding of effective motivational and instructional designs in education and contribute to the increasing reproducibility of such research. The resulting project artifacts will help to build the foundation for collaborative open research practices well fitted to educational research and continuous improvement practice.

This work was partially supported by the National Science Foundation (#2209819) project Collaborative Research: Frameworks for Intelligent Adaptive Experimentation: Enhancing and Tailoring Digital Education, the Natural Sciences and Engineering Research Council of Canada (NSERC) (#RGPIN-2019-06968), as well as by the Office of Naval Research (ONR) (#N00014-21-1-2576).

## REFERENCES

Amiel, T., & Reeves, T. C. (2008). Design-based research and educational technology: Rethinking technology and the research agenda. *Journal of Educational Technology & Society*, *11*(4), 29–40.

Bouneffouf, D., Rish, I., & Aggarwal, C. (2020). Survey on Applications of Multi-Armed and Contextual Bandits. *2020 IEEE Congress on Evolutionary Computation (CEC)*, 1–8. https://doi.org/10.1109/CEC48606.2020.9185782

Bryan, C. J., Tipton, E., & Yeager, D. S. (2021). Behavioural science is unlikely to change the world without a heterogeneity revolution. *Nature Human Behaviour*, *5*(8), 980–989. https://doi.org/10.1038/s41562-021-01143-3

Chen, P., Teo, D. W. H., Foo, D. X. Y., Derry, H. A., Hayward, B. T., Schulz, K. W., Hayward, C., McKay, T. A., & Ong, D. C. (2022). Real-world effectiveness of a social-psychological intervention translated from controlled trials to classrooms. *Npj Science of Learning*, *7*(1), 20. https://doi.org/10.1038/s41539-022-00135-w

Deng, A., Dmitriev, P., Gupta, S., Kohavi, R., Raff, P., & Vermeer, L. (2017). A/B Testing at Scale: Accelerating Software Innovation. *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1395–1397. https://doi.org/10.1145/3077136.3082060

Fancsali, S., Murphy, A., & Ritter, S. (2022). "Closing the Loop" in Educational Data Science with an Open Source Architecture for Large-Scale Field Trials. In A. Mitrovic & N. Bosch (Eds.), *Proceedings of the 15th International Conference on Educational Data Mining* (pp. 834–838). International Educational Data Mining Society. https://doi.org/10.5281/zenodo.6852930

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://doi.org/10.1007/s40593-014-0024-x

Kale, A., Lee, S., Goan, T., Tipton, E., & Hullman, J. (2023). MetaExplorer: Facilitating Reasoning with Epistemic Uncertainty in Meta-analysis. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–14. https://doi.org/10.1145/3544548.3580869

Kizilcec, R. F., Reich, J., Yeomans, M., Dann, C., Brunskill, E., Lopez, G., Turkay, S., Williams, J. J., & Tingley, D. (2020). Scaling up behavioral science interventions in online education. *Proceedings of the National Academy of Sciences*, *117*(26), 14900–14905. https://doi.org/10.1073/pnas.1921417117

Motz, B. A., Carvalho, P. F., De Leeuw, J. R., & Goldstone, R. L. (2018). Embedding Experiments: Staking Causal Inference in Authentic Educational Contexts. *Journal of Learning Analytics*, *5*(2). https://doi.org/10.18608/jla.2018.52.4

Musabirov, I. (2022). Challenges and Opportunities of Infrastructure-Enabled Experimental Research in Computer Science Education. *Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 2*, 1265–1265. https://doi.org/10.1145/3545947.3573242

Reza, M., Kim, J., Bhattacharjee, A., Rafferty, A. N., & Williams, J. J. (2021). The MOOClet Framework: Unifying Experimentation, Dynamic Improvement, and Personalization in Online Courses. *Proceedings of the Eighth ACM Conference on Learning@ Scale*, 15–26.

Sankaranarayanan, S., Harrison, C., Kumari, S., Larson, K., Lemiesz, R., Mesa, N., Mitts, R., Verma, E., & Zimmaro, D. (2023). Building an infrastructure for A/B experiments at scale: The challenges, opportunities, and lessons for the learning analytics community. *Companion Proceedings 13 Th International Conference on Learning Analytics & Knowledge (LAK23)*, 37–40. https://www.amazon.science/publications/building-an-infrastructure-for-a-b-experiments-at-scale-the-challenges-opportunities-and-lessons-for-the-learning-analytics-community

Stamper, J. C., Lomas, D., Ching, D., Ritter, S., Koedinger, & K.R., J., Steinhart. (2012). The rise of the super experiment. *Proceedings of the 5th International Conference on Educational Data Mining*, 196–199.

# Self-regulated learning and off-task thoughts in online learning: Investigating the interaction during learning from videos

**Daniel Ebbert**
University of South Australia
daniel.ebbert@mymail.unisa.edu.au

**ABSTRACT**: Self-regulated learning is critical for success in online learning. However, students inevitably experience off-task thoughts (mind wandering) that can disrupt learning. Although these two factors have been studied independently, the relationship between self-regulated learning and off-task thoughts has not been studied extensively. This research explores the relationship between self-regulated learning and off-task thoughts while learning online from a video. A mixed methods approach combines meta-analysis, a case study, an experiment, and comparative analysis to investigate off-task thought frequency and its influence on self-regulation processes. A conceptual paper will present a model of how off-task thoughts may prompt reactive self-regulation during learning. Meta-analysis will synthesize the occurrence and impact of task-related interference. A naturalistic case study and controlled experiment will gather self-caught thought reports during actual and simulated video learning. Comparing results will assess generalizability across contexts. This research will provide theoretical and empirical insights into the relationship between off-task thoughts and self-regulated learning when learning from videos.

**Keywords**: self-regulated learning, mind wandering, metacognition, meta-awareness, off-task thought

## 1 INTRODUCTION

When attempting to learn, it is common for students to think about something unrelated. A student will think about something else about 30% of the time during educational activities (Wong et al., 2022). This is called task-unrelated thought or, more generally, off-task thought, and it is inevitable. Therefore, off-task thoughts must be considered when studying how students learn. How students adapt their learning to the current situation, including distractions, is part of self-regulated learning (Panadero, 2017). However, distractions, such as off-task thoughts, can occur while executing the learning process, and students then need to adapt to this distraction on the fly.

## 2 BACKGROUND

### 2.1 Off-task thoughts

Off-task thoughts can be conceptualized along the two dimensions of *stimulus-dependency* and *task-relatedness* (Stawarczyk et al., 2011). For the context of this research project, only stimulus-independent thoughts are of interest, and these can be either task-unrelated or task-related. *Stimulus-independent and task-unrelated* thoughts are called mind wandering, and *stimulus-independent and task-related* thoughts are called task-related interference.

Based on a recent meta-analysis, mind wandering occurs about 30% of the time during educational activities and has a negative relationship with learning outcomes (Wong et al., 2022). A comparable meta-analysis on the frequency and effect on learning outcomes of task-related interference has not been conducted yet. When students are learning, they will encounter both types of off-task thoughts and adjust their learning depending on the type of off-task thought they encounter.

Because of the negative effect of mind wandering on learning outcomes, there have been various lab studies whose authors attempted to reduce how often the learners are mind wandering. In the context of learning from videos, this has been achieved using interpolated testing (Jing et al., 2016; Szpunar et al., 2013; Welhaf et al., 2022). Other types of learning activities could be incorporated into a video. For example, generative activities such as self-explanations positively affect learning outcomes (Fiorella & Mayer, 2015), but it is unknown which effect they will have on of off-task thought.

It is also possible that students realize they were mind wandering; this is called meta-awareness (Schooler et al., 2011). This is a form of metacognitive monitoring (Schooler & Smallwood, 2009). When students engage in metacognitive monitoring, they are actively thinking about what they were just thinking about, and when they do this, they could become aware that they were just off-task. The information gained about their thoughts can then be used as the basis for adjusting ones' thoughts. This is called metacognitive control. A student then studies until they engage in metacognitive monitoring again, which can trigger metacognitive control. This metacognitive monitoring and control cycle forms the basis of self-regulated learning (Winne & Azevedo, 2022).

### 2.2 Self-regulated learning

Self-regulated learning is a framework for understanding the emotional, motivational, and cognitive aspects of learning (Panadero, 2017). The self-regulated learning model chosen to underpin this research is the COPES model (Winne & Hadwin, 1998) because it describes how metacognition is part of self-regulated learning and how students adapt their learning process to the current task. Following Winne and Hadwin (1998), self-regulated learning occurs across four interconnected stages and during the third phase, while the students enact study tactics and strategies, they frequently switch between cognition and metacognitive monitoring (Winne, 2011). During this phase, students are theoretically most likely to realize they were having off-task thoughts and then adapt their learning behavior based on this realization. It is this phase of self-regulated learning that is being explored in this research project and so far there is no model describing how off-task thoughts influence self-regulated learning.

## 3 RESEARCH APPROACH

This research project consists of two parts. The first part is theory development, complemented by a meta-analysis. Together, these inform the second part, exploring self-regulated learning and off-task thoughts during learning from videos.

### 3.1 Theory development and meta-analysis

Building upon the existing literature, a model will be developed that outlines how off-task thoughts influence learning and how students might react to realizing their off-task thoughts. The model will be based on the COPES model of self-regulated learning, theories on off-task thought, and the concept of metacognition. This model will be presented in a conceptual paper. Part of this model stipulates that how a self-regulated learner reacts to realizing they were off-task depends on the type of off-task thought they had. While the frequency and relationship with learning outcomes are already known for mind wandering (Wong et al., 2022), this is not the case for task-related interference, which motivates the first two research questions.

- **RQ1:** How often does task-related interference occur during learning?
- **RQ2:** How strong is the relationship between task-related interference and learning outcome?

The developed model will serve as a theoretical basis, enriched by the meta-analysis on task-related interference and complemented by the meta-analysis on task-unrelated thought by Wong et al. (2022). Together, these components will describe how frequently students face each type of off-task thought and thus need to react to it. In the following, the context of video-based learning has been chosen to explore these dynamics in detail and assess their practical implications.

### 3.2 Exploring self-regulated learning and off-task thoughts during video learning

The theoretical assumption of mutual influence between self-regulated learning and off-task thoughts led to the overarching question, "*What is the 2-way relationship between self-regulated learning and off-task thoughts in video-based learning?*" The overarching research question has been broken down into specific research questions.

- **RQ3:** How does self-regulated learning influence off-task thoughts when learning from a video in a naturalistic environment?
- **RQ4:** How does self-explanation during video watching influence off-task thoughts compared to interpolated testing?
- **RQ5:** Is the relationship between self-regulated learning and off-task thought frequency consistent or different when comparing naturalistic (case study) and controlled (experiment) settings?

The limitation that most off-task thought and learning research has been conducted in lab environments motivated RQ3; that the reduction of off-task thought during learning from videos has mostly been attempted using interpolated testing motivated RQ4; and concerns about the applicability of lab-based research results to a naturalistic setting motivated RQ5.

## 4 METHODOLOGY

A combination of methods will address the previously described research questions. These are a meta-analysis, a case study, and an experiment. The data from the case study and the experiment will then be combined to analyze and compare the frequency of the off-task thoughts and assess the potential impact of self-regulated learning on off-task thoughts. In both the case study and experiment, self-caught free-text thought reports are used, and the participants are asked to answer subscales[1] from the self-regulation for learning online (SRL-O) questionnaire (Broadbent et al., 2022).

### 4.1 Meta-analysis

The research questions one and two will be approached using a meta-analysis. Systematically, the existing literature on task-related interference will be searched and screened. Afterwards, the frequency and effect size for the relationship between task-related interference and learning outcomes will be extracted and included in a meta-analysis. This information will reveal how often task-related interference occurs and how strongly it affects learning outcomes. This study has been pre-registered[2].

### 4.2 Case study

The third research question will be approached using an exploratory case study across multiple courses at the University of South Australia. A significant limitation of many studies on the relationship of off-task thought and learning is that these studies took place in controlled laboratory environments. For the context of studying off-task thoughts while learning from a video, this also meant that the learners did not have a chance to react to the realization of their off-task thoughts, although they might have wanted to. This study aims to overcome this limitation by asking students to watch videos that are part of their courses and report their off-task thoughts when they realize them. Furthermore, unlike other studies on off-task thoughts, students can interact with the video player while learning from videos. They, therefore, could react to the realization of their off-task thoughts. The resulting trace data consisting of thought reports and video interaction data can be analyzed using sequential pattern mining. The participating students will watch multiple videos during their participation, leading to more sequences that can be analyzed. Specifically, the study aims to involve approximately 100 students, expecting to generate around 200 distinct sequences for analysis. In addition to measuring the students' self-caught off-task thoughts, they will be asked to answer SRL-O subscales at the beginning of their participation. Based on the answer to the SRL-O subscales and their thought reports, a multilevel model can be constructed to model the relationship between self-regulated learning and off-task thought. The results from this study will provide insight into how often off-task thoughts occur in a naturalistic setting and if the students react if they realize they were off task by, for example, rewinding the video.

### 4.3 Experiment

The fourth research question will be addressed using an experiment. An experiment will be designed in which the effect of interpolated testing and writing self-explanations of the video content on the self-reported frequency of off-task thoughts is compared between the two experimental conditions and with a control group. This experiment will consist of SRL-O subscales, a pre-test, watching the video and reporting off-task thoughts (self-caught), a filler task, and a post-test. While the participants

[1] Subscales used in both: Metacognition, Online Effort Regulation, and Online Task Strategies. Subscales additionally used in the case study: Online Planning and Time Management, and Online Study Environment.

[2] https://osf.io/znhfy

watch the video, they are instructed to report any off-task thoughts they realize. Additionally, each time the participants are interrupted for a learning activity (interpolated testing or self-explanation), they are asked to write a thought report. The results from this study will provide insights into which learning activity (interpolated testing or self-explanations) will lead to better learning outcomes and if the realization frequency of off-task thoughts differs between the two experimental conditions and a control group. This study has been pre-registered[3].

### 4.4 Comparison of off-task task thought frequency between case study and experiment

Once the case study and the experiment are completed, research question five can be answered. The data from the case study will be combined with the control group's data from the experiment. Combining the data, the frequency, and types of off-task thoughts between the two studies can be compared based on the participants' SRL-O questionnaire scores. The resulting data can provide insight into students' self-regulation in different contexts and might describe that students with a similar score on the SRL-O subscales have a similar or different frequency of self-caught off-task thought frequency in the case study than in the experiment.

## 5 STATUS

The status of this research project is that the conceptual paper is being written in cooperation with Caitlins Mills from the University of Minnesota and Phil Winne from Simon Fraser University. The meta-analysis is conducted with Andrew Zamecnik from the University of South Australia. The database searches for this meta-analysis have already been conducted, and the resulting sources are being screened. For the case study, the data collection is in progress, and the data collection for the experiment is expected to be completed before LAK24.

## 6 ETHICAL CONSIDERATIONS

Ethical approval has been obtained from the Human Research Ethics Committee of the University of South Australia. The personal information obtained during the case study is de-identified before data analysis, and only anonymous data is collected for the experiment. Only data from participants who provided informed consent will be used in both cases.

## 7 CONTRIBUTION

Overall, this research will contribute to the existing literature by investigating the relationship between off-task thoughts and self-regulated learning and how students might influence their off-task thought frequency after realizing they were off-task. Additionally, this research provides evidence of learning and expands learning theories that can be used to inform automated interventions to reduce mind wandering and thereby improve learning outcomes.

[3] https://osf.io/4dg5u

## REFERENCES

Broadbent, J., Panadero, E., Lodge, J. M., & Fuller-Tyszkiewicz, M. (2022). The self-regulation for learning online (SRL-O) questionnaire. *Metacognition and Learning*. https://doi.org/10.1007/s11409-022-09319-6

Fiorella, L., & Mayer, R. E. (2015). *Learning as a Generative Activity: Eight Learning Strategies that Promote Understanding* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781107707085

Jing, H. G., Szpunar, K. K., & Schacter, D. L. (2016). Interpolated testing influences focused attention and improves integration of information during a video-recorded lecture. *Journal of Experimental Psychology: Applied*, *22*(3), 305–318. https://doi.org/10.1037/xap0000087

Panadero, E. (2017). A Review of Self-regulated Learning: Six Models and Four Directions for Research. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00422

Schooler, J. W., & Smallwood, J. (2009). Metacognition. In T. Bayne, A. Cleeremans, & P. Wilken (Eds.), *The Oxford Companion to Consciousness* (1st ed., pp. 438–442). Oxford University Press.

Schooler, J. W., Smallwood, J., Christoff, K., Handy, T. C., Reichle, E. D., & Sayette, M. A. (2011). Meta-awareness, perceptual decoupling and the wandering mind. *Trends in Cognitive Sciences*, *15*(7), 319–326. https://doi.org/10.1016/j.tics.2011.05.006

Stawarczyk, D., Majerus, S., Maj, M., Van der Linden, M., & D'Argembeau, A. (2011). Mind-wandering: Phenomenology and function as assessed with a novel experience sampling method. *Acta Psychologica*, *136*(3), 370–381. https://doi.org/10.1016/j.actpsy.2011.01.002

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences*, *110*(16), 6313–6317. https://doi.org/10.1073/pnas.1221764110

Welhaf, M. S., Phillips, N. E., Smeekens, B. A., Miyake, A., & Kane, M. J. (2022). Interpolated testing and content pretesting as interventions to reduce task-unrelated thoughts during a video lecture. *Cognitive Research: Principles and Implications*, *7*(1), 26. https://doi.org/10.1186/s41235-022-00372-y

Winne, P. H. (2011). A Cognitive and Metacognitive Analysis of Self-Regulated Learning. In D. H. Schunk & B. Zimmerman (Eds.), *Handbook of Self-Regulation of Learning and Performance* (1st ed., pp. 15–32). Routledge.

Winne, P. H., & Azevedo, R. (2022). Metacognition and Self-Regulated Learning. In R. K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (3rd ed., pp. 93–113). Cambridge University Press. https://doi.org/10.1017/9781108888295.007

Winne, P. H., & Hadwin, A. F. (1998). Studying as Self-Regulated Learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in Educational Theory and Practice* (1st ed., pp. 277–304). Routledge.

Wong, A. Y., Smith, S. L., McGrath, C. A., Flynn, L. E., & Mills, C. (2022). Task-unrelated thought during educational activities: A meta-analysis of its occurrence and relationship with learning. *Contemporary Educational Psychology*, *71*, 102098. https://doi.org/10.1016/j.cedpsych.2022.102098

# Learning Analytics of Theory-Practice Integration in Vocational Education and Training

**Wanruo Shi**
Tsinghua University
swr22@mails.tsinghua.edu.cn

**ABSTRACT**: Integrating theoretical and practical learning remains challenging in developing students' vocational competence and professional expertise. Learning analytics (LA) holds the potential to offer valuable insights into student integrative learning in vocational education and training (VET), allowing for better student support and targeted pedagogical interventions. Despite significant strides in LA research utilizing trace data to analyze student behaviors and predict academic outcomes over the past decade, scant attention has been devoted to investigating the nuanced interplay between theoretical and practical learning within the context of VET. This doctoral research aims to utilize LA to address the challenges of comprehending the links and gaps between theoretical and practical learning. This will be achieved through the integration of diverse learning data sources, including but not limited to learning management systems, placement assessments, and practical training platforms. Furthermore, this research endeavours to broaden the scope of LA investigations to encompass VET, an area that has hitherto received comparatively inadequate attention in the existing body of research. The anticipated outcomes of this research include enhanced LA in VET, contributing to usable guidance and insights on integrative learning of theory and practice.

**Keywords**: Learning analytics; Theory-Practice Integration; Vocational education and training (VET); Multiple data traces

## 1 INTRODUCTION

Vocational education and training (VET) involve the interconnected realms of theoretical and practical learning. For VET students, both theoretical and practical learning are of great importance to acquire professional knowledge and skills, and to develop their occupational competencies. Theory-practice integration is regarded as one of the fundamental learning features of VET(Hiim, 2017). The distinction between **theoretical learning** and **practical learning** lies in the forms of knowledge. Theoretical knowledge, in contrast to applied knowledge, is characterized by its decontextualized and general nature (Van De Ven & Johnson, 2006). Simultaneously, vocational schools and workplaces are the two main learning scenarios of VET, where theoretical and practical learning should be integrated (Schaap et al., 2012).

However, integrating theoretical and practical learning in VET remains challenging. The theory-practice gap has long been discussed and recognized as one of the critical concerns in developing professional expertise and vocational competence (Dadgaran et al., 2012). Qualitative investigations through individual and group interviews show that students can recognise the importance of integrating theoretical and practical learning in VET (Baartman et al., 2018). Nevertheless, stimulating students' capacity to achieve the integration are often implicit and not always guaranteed within the VET curriculum. In the pursuit of enhancing the integration of theoretical and practical learning, professional culture creation and curriculum content reformation are proposed solutions (Saifan et al., 2021), along with learning strategies and teaching approach such as learning by reflection and the jigsaw strategy (O'Leary et al., 2015). Despite the scholarly attention on the theory-practice integration in VET, primarily relying on qualitative and subjective evidence through case studies, interviews, and action research, there is limited data-based and objective evidence to comprehensively elucidate the integrative process.

With recent technology advances and rises in data-intensive scientific discoveries(Tolle et al., 2011), one area that can aid in understanding the integration of theoretical and practice learning is Learning Analytics (LA). LA contributes to predicting learning outcomes, providing learning feedback and early interventions, and promoting personalized learning. Within this domain, multimodal learning analytics (MMLA) techniques involving sensor technology, have the potential to provide insights into the learning processes in VET settings. Such insights could hold the promise of facilitating enhanced student support and targeted pedagogical interventions.

This doctoral research endeavors to confront the challenges in comprehending the links and gaps between theoretical and practical learning through utilizing multiple data traces in VET. These diverse data traces, derived from the integrative learning process, encompass data gleaned from learning management systems, practical training platforms, placement assessments, recorded interactions, and group interviews, among other sources. The amalgamation of these various data streams holds the potential to provide distinctive insights into the intricate interplay and integration of theoretical and practical learning within the context of VET.

# 2 BACKGROUND

## 2.1 The Integrative Learning Nature in VET

The integrative learning of theory and practice has been widely acknowledged as a cornerstone of VET, where school-based learning and workplace learning are closely convergent (Orozco et al., 2019). The expectation for students is to cultivate occupational competencies by seamlessly integrating theoretical understanding with hands-on experience. Vocational schools play a dual role in this process, serving as environments for acquiring theoretical knowledge and engaging in meaningful peer interactions, while also providing simulated or virtual practical training to enable the development of professional skills and the establishment of initial professional identities (Collin & Tynjälä, 2003). Simultaneously, workplace learning complements practical training by immersing students in authentic work environments, allowing them to derive theoretical insights through reflection on their practical experiences as well(Griffiths & Guile, 2003). Consequently, the integration of theory and practice unfolds in both vocational schools and workplaces. The intricate nature of this integrative learning process in VET necessitates a comprehensive exploration with empirical data evidence.

## 2.2 LA in Integrative Learning Scenarios

Literature investigations into LA within VET settings suggest that the full potential of LA remains underutilized in vocational education (Gedrimiene et al., 2020), particularly in unraveling the intricate relationship between theoretical and practical learning. The multifaceted nature of vocational learning environments introduces challenges arising from the complexity, diversity, and authenticity of data dispersed across various platforms and sites, complicating the collection and analysis of learning data (Santamaría-Bonfil et al., 2021). This section endeavours to consolidate insights from diverse perspectives by examining LA research not confined solely to VET settings. Aiming at providing a comprehensive examination of LA in integrative learning scenarios, the review broadens its scope to encompass LA applications in simulated learning environments, virtual laboratories, project/problem-based learning (PBL) contexts, and professional education settings.

Within the framework of the PBL-LA model, a four-layer structure encompassing data, analytics, pedagogical, and ICT layers has been developed and implemented in ten online courses (Zotou et al., 2020). Social network analysis has been leveraged to monitor online interactions in PBL, contributing to the understanding of learners' collaborative dynamics(Saqr & Alamro, 2019). Additionally, MMLA also plays a role in learning performance prediction in integrative learning scenarios of PBL. For instance, procedural data were gathered using multiple sensors to identify key indicators that predict students' performance in open-ended complex learning tasks (Spikol et al., 2018).

Furthermore, the application of LA in teacher education and medical education has provided valuable insights for understanding the integrative learning process. Focusing on developing teachers' technological pedagogical content knowledge (TPACK), multiple data traces have been used to assess teacher's self-regulated learning in planning lessons by combining data from computer logs and think-aloud data. Results show that MMLA contributes to explaining a significant variance in TPACK performance(Huang et al., 2023). Visual learning analytics based on data evidence extraction, supports classroom discourse analysis in teachers' professional learning and development as well (Chen, 2019). In the realm of medical education, there is also evidence that nursing students achieve theory-practice integration through immersive simulation. MMLA with sensor or wearable technology also holds the potential to assess the integrative learning in procedural medicine(Mohamadipanah et al., 2021).

# 3 RESEARCH QUESTIONS

Building upon the insights gathered from the literature review above, preliminary investigations into LA have been undertaken within specific integrative learning contexts, including teacher education, medical education, and PBL scenarios. However, there exists a need for more extensive and nuanced LA research within the realm of VET, specifically to comprehensively explore the intricate dynamics between theoretical and practical learning. This doctoral research endeavours to contribute to this gap in the existing body of knowledge by implementing LA with multiple data traces. The primary aim is to gain a deeper understanding of the links and gaps between theoretical learning and practical training within the VET context. To address this overarching research objective, three specific research questions are outlined as follows:

***RQ1:*** *How is LA used to measure the learning processes in VET?*

**- RQ 1.1:** How is LA used to measure the theoretical learning processes in VET?

**- RQ 1.2:** How is LA used to measure the practical learning processes in VET?

***RQ2:*** *Can LA be used to identify links and gaps between theoretical learning and practical learning?*

**- RQ 2.1:** Can we use the theoretical learning traces to predict the practical learning performance?

**- RQ 2.2:** Can we use the practical assessment data to identify the relationship among professional knowledge, practice, and engagement?

**- RQ 2.3:** Can we use the holistic learning process data to identify the gap between theoretical learning and practical learning in VET?

***RQ3:*** *What constructs should be included in the framework for LA in VET?*

# 4 METHODOLOGY

The research methodology for this study is structured into three phases, each aligned with a specific research question. Figure 1 visually represents the interconnectedness of these phases, emphasizing the close relationship between RQ2 and RQ3.

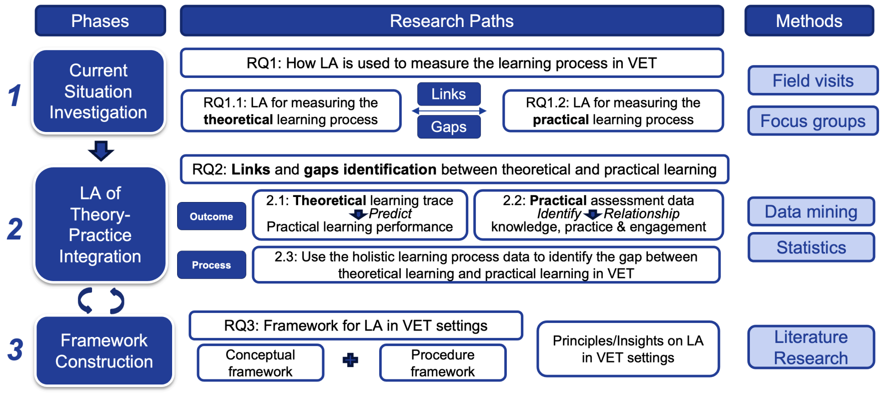



**Figure 1 The research map**

To address RQ1, the first phase involves the field study in nine vocational colleges in China. Through the field study, nine focus group interviews and site visits were conducted to reveal how LA is used to measure the learning processes in VET, including both theoretical and practical learning scenarios. Further, the potential links and gaps were initially analyzed between theoretical and practical learning. Additionally, the phase aids in the exploration of data sources and the identification of challenges that will inform subsequent phases of the research.

Aligning with RQ2, the second phase focuses on using LA to identify the interplay of theoretical and practical learning in VET. Integrative learning outcomes are analyzed by utilizing students' theoretical learning process data to predict their practical learning performance. Moreover, placement assessment data in Initial Teacher Education programs is scrutinized to understand the relationships among professional knowledge, practice, and engagement. This phase also investigates the integrative learning process by collecting and analyzing holistic learning process data in VET. The goal is to identify gaps between theoretical and practical learning processes, providing a comprehensive understanding of the dynamics between the two in VET.

The final phase of the research is dedicated to addressing RQ3 and involves the construction of conceptual and procedural frameworks for LA in VET. Building upon the findings from the preceding phases, this stage aims to illustrate how Learning Analytics can be effectively applied in integrative learning scenarios to bridge the gap between theoretical and practical learning. The outcomes of this phase include the development of framework and guiding principles on the implementation of LA in VET, aiming to inform practitioners and educators on better practices in integrative learning of VET.

## 5 CURRENT PROGRESS

The first research phase for investigating the current situation of LA in VET has been completed through nine focus group interviews and field visits. The research is currently progressing through the preliminary stage of Phase 2. To answer RQ2.1, a pilot study was conducted with students' log data and practical assessment data in an Initial Math Teacher Education Program in China. According to the pilot, theoretical learning traces can predict students' practical learning performance in the integrative course. Also, procedural data performs better than aggregated data in the prediction models.

As part of answering RQ2.2, placement data from Initial Teacher Education Programs in Australia has been collected. The ongoing efforts involve exploring the relationships among professional knowledge, practice, and engagement using a Structural Equation Model (SEM). The current progress reflects a robust foundation for further exploration and analysis in the subsequent stages of the research.

## 6 CONTRIBUTION

***Theoretical contribution***: While LA research has already delved into using trace data to analyze students' behaviors and predict academic achievement over the last decade, limited research has focused on understanding the integrative process of theory and practice. This research intends to bridge the theoretical and practical learning gap by conducting LA research. It represents a valuable contribution to the field of LA in VET settings.

***Practical contribution***: Bridging the theory-practice gap remains a persistent challenge in VET contexts. To tackle this challenge, this doctoral research employs LA that draws from both theoretical and practical learning traces. Through this exploration, we aim to develop the framework and guiding principles of implementing LA in VET, informing practitioners and educators on better practices in integrative learning of VET.

## REFERENCES


Allan, H. T. (2011). Using psychodynamic small group work in nurse education: Closing the theory–practice gap? Nurse Education Today, 31(5), 521–524. https://doi.org/10.1016/J.NEDT.2010.09.006

Avella, J. T., Kebritchi, M., Nunn, S. G., & Kanai, T. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. Journal of Asynchronous Learning Network, 20(2). https://doi.org/10.24059/olj.v20i2.790

Baartman, L. K. J., Kilbrink, N., & de Bruijn, E. (2018). VET students' integration of knowledge engaged with in school-based and workplace-based learning environments in the Netherlands. Journal of Education and Work, 31(2). https://doi.org/10.1080/13639080.2018.1433821

Beckett, D. (2000). Making Workplace Learning Explicit: an epistemology of practice for the whole person. Westminster Studies in Education, 23(1), 41–53. https://doi.org/10.1080/0140672000230105

Berman, N. B., & Artino, A. R. (2018). Development and initial validation of an online engagement metric using virtual patients. BMC Medical Education, 18(1). https://doi.org/10.1186/s12909-018-1322-z

Chen, G. (2019). Visual learning analytics to support classroom discourse analysis for teacher professional learning and development. In The Routledge International Handbook of Research on Dialogic Education (pp. 167–181). Routledge. https://doi.org/10.4324/9780429441677-15

Collin, K., & Tynjälä, P. (2003). Integrating theory and practice? Employees' and students' experiences of learning at work. Journal of Workplace Learning, 15(7/8), 338–344. https://doi.org/10.1108/13665620310504828

Dadgaran, I., Parvizy, S., & Peyrovi, H. (2012). A Global Issue in Nursing Students' Clinical Learning: The Theory–Practice Gap. Procedia - Social and Behavioral Sciences, 47, 1713–1718. https://doi.org/10.1016/j.sbspro.2012.06.888

Eiríksdóttir, E. (2020). Program coherence and integration of school-and work-based learning in the icelandic dual vocational education and training (Vet) system. Education Sciences, 10(11). https://doi.org/10.3390/educsci10110314

Gedrimiene, E., Silvola, A., Pursiainen, J., Rusanen, J., & Muukkonen, H. (2020). Learning Analytics in Education: Literature Review and Case Examples From Vocational Education. Scandinavian Journal of Educational Research, 64(7), 1105–1119. https://doi.org/10.1080/00313831.2019.1649718

Griffiths, T., & Guile, D. (2003). A Connective Model of Learning: The Implications for Work Process Knowledge. European Educational Research Journal, 2(1). https://doi.org/10.2304/eerj.2003.2.1.10

Hiim, H. (2017). Ensuring curriculum relevance in vocational education and training: Epistemological perspectives in a curriculum research project. International Journal for Research in Vocational Education and Training, 4(1). https://doi.org/10.13152/IJRVET.4.1.1

Hossain, Z., Bumbacher, E., Brauneis, A., Diaz, M., Saltarelli, A., Blikstein, P., & Riedel-Kruse, I. H. (2018). Design Guidelines and Empirical Case Study for Scaling Authentic Inquiry-based Science Learning via Open Online Courses and Interactive Biology Cloud Labs. International Journal of Artificial Intelligence in Education, 28(4). https://doi.org/10.1007/s40593-017-0150-3

Huang, L., Doleck, T., Chen, B., Huang, X., Tan, C., Lajoie, S. P., & Wang, M. (2023). Multimodal learning analytics for assessing teachers' self-regulated learning in planning technology-integrated lessons in a computer-based environment. Education and Information Technologies. https://doi.org/10.1007/s10639-023-11804-7

Mohamadipanah, H., Wise, B., Witt, A., Goll, C., Yang, S., Perumalla, C., Huemer, K., Kearse, L. D., & Pugh, C. (2021). Performance assessment using sensor technology. In Journal of Surgical Oncology (Vol. 124, Issue 2). https://doi.org/10.1002/jso.26519

O'Leary, N., Wattison, N., Edwards, T., & Bryan, K. (2015). Closing the theory–practice gap. European Physical Education Review, 21(2), 176–194. https://doi.org/10.1177/1356336X14555300

Orozco, M., Gijbels, D., & Timmerman, C. (2019). Empirical Conceptualisation of Integrative Learning. A Focus on Theory-Practice Integration in Technical Vocational Education and Training. Vocations and Learning, 12(3). https://doi.org/10.1007/s12186-019-09223-2

Pugh, C. M., Ghazi, A., Stefanidis, D., Schwaitzberg, S. D., Martino, M. A., & Levy, J. S. (2020). How Wearable Technology Can Facilitate AI Analysis of Surgical Videos. Annals of Surgery Open, 1(2). https://doi.org/10.1097/as9.0000000000000011

Saifan, A., Devadas, B., Daradkeh, F., Abdel-Fattah, H., Aljabery, M., & Michael, L. M. (2021). Solutions to bridge the theory-practice gap in nursing education in the UAE: a qualitative study. BMC Medical Education, 21(1), 490. https://doi.org/10.1186/s12909-021-02919-x

Santamaría-Bonfil, G., Díaz-Rodríguez, H. D., Arroyo-Figueroa, G., & Batres, R. (2021). Knowledge modelling for ill-defined domains using learning analytics: Lineworkers case. In Advances in Intelligent Systems and Computing: Vol. 1266 AISC. https://doi.org/10.1007/978-3-030-57799-5_42

Saqr, M., & Alamro, A. (2019). The role of social network analysis as a learning analytics tool in online problem based learning. BMC Medical Education, 19(1). https://doi.org/10.1186/s12909-019-1599-6

Schaap, H., Baartman, L., & de Bruijn, E. (2012). Students' Learning Processes during School-Based Learning and Workplace Learning in Vocational Education: A Review. Vocations and Learning, 5(2), 99–117. https://doi.org/10.1007/s12186-011-9069-2

Spikol, D., Ruffaldi, E., Dabisias, G., & Cukurova, M. (2018). Supervised machine learning in multimodal learning analytics for estimating success in project-based learning. Journal of Computer Assisted Learning, 34(4). https://doi.org/10.1111/jcal.12263

Tolle, K. M., Tansley, D. S. W., & Hey, A. J. G. (2011). The fourth Paradigm: Data-intensive scientific discovery. Proceedings of the IEEE, 99(8). https://doi.org/10.1109/JPROC.2011.2155130

Van De Ven, A. H., & Johnson, P. E. (2006). Knowledge for theory and practice. Academy of Management Review, 31(4). https://doi.org/10.5465/amr.2006.22527385

# Examining self-regulated learning through regular reflective practice

**Nazanin Rezazadeh Mottaghi**
University of South Australia
Rezny004@mymail.unisa.edu.au

**ABSTRACT:** A key challenge for tertiary education institutions lies in supporting students to learn independently. To overcome this challenge, students require the ability to plan, monitor and appraise their learning effectively in relation to institutional expectations (Vosniadou, 2020). Self-regulated learning skills can help students to gain abilities to regulate aspects of their thinking, motivation and behavior throughout their learning processes (Panadero, Broadbent, Boud, & Lodge, 2018). Using reflective writing could be regarded as an approach which supports students to evaluate their work independently, monitor their plans to gain goals and improve SRL skills (Jung & Wise, 2020). There is little work in this field that investigates how regular self-reflective writing on assignment task can improve students' SRL skills and ultimately improve their academic achievement. Therefore, this study aims to address this gap to investigate the effect of regular reflective writing on their SRL skills by examining students' reflective writing tasks and associated assessment marks through automated text analytics tools and statistical analysis methods in a range of courses from different disciplines and year levels.

**Keywords**: Self-regulated learning, Reflective writing, Metacognition, Judgement of learning

## 1 INTRODUCTION

University education demands students to transition from structured environments to autonomous learning. The shift requires rapid development of independent learning skills crucial for navigating the unstructured nature of higher education (Vosniadou, 2020). Many new students, however, struggle with this transition (Boud, 2010), hindering their ability to grasp self-knowledge and learning processes critical for future education. This struggle often arises due to inadequate preparation and challenges faced in transitioning from secondary to higher education (Zimmerman, 2002). Self-regulated learning (SRL), a framework encompassing motivation, metacognition, and cognition, becomes pivotal in addressing these challenges (Chen & Cheng, 2020). SRL empowers learners to transform mental abilities into academic performance skills (Zimmerman, 2008), aligning with the educational goal of enabling students to regulate their own learning effectively (Delors, 1996). Hence, fostering skills and strategies for self-regulated learning is essential in university education to equip students for autonomous learning.

The doctoral research focuses on enhancing students' self-regulatory skills in higher education through self-reflection. It explores employing text analytics due to time constraints in analyzing reflective writing. The background covers self-regulated learning, self-reflection, and students' metacognition to frame the study.

## 2 BACKGROUND

### 2.1 Self-regulated learning

Self-regulated learning can be shown as "proactive processes" (Zimmerman, 1990). These processes enable learners to select their goals, set their plans, identify strategies and monitor the effectiveness of their learning. The aim of these proactive processes is to support learners to display the capabilities of perseverance, personal initiative and adaptive skills from motivational feelings and beliefs and metacognitive strategies in order to improve their academic achievements (Zimmerman, 1990). While there are multiple models of SRL including Zimmerman (Zimmerman, 1990; 2002); Boekaerts (Boekaerts, 1996); Winne and Hadwin (Winne & Hadwin, 1998); Pintrich (Pintrich, Wolters, & Baxter, 2000); Efklides (2011) but the main components of the models are similar. There is an agreement between the models that the SRL process is a recursive cycle and has certain phases such as: preparation, performance and appraisal (self-reflection) as illustrated in Figure 1 (Storie, 2021). During these phases, learners typically undertake a variety of activities such as selecting their goals, setting their plans and choosing learning strategies, and continuously monitoring their progress against their goals and then reflecting on their outcome in order to improve their learning in the future (Storie, 2021).

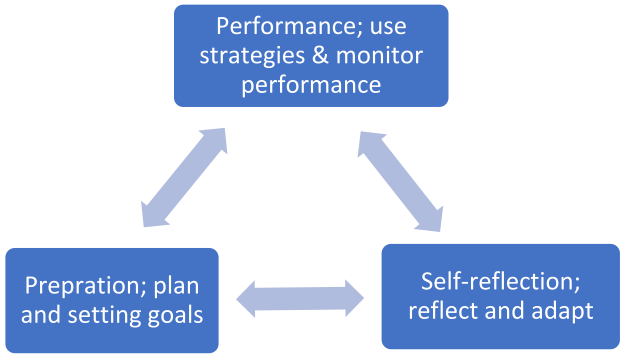



Figure 1: Different phases of self-regulated learning adapted from Storie 2021

This doctoral research is grounded in the COPES model proposed by Winne and Hadwin (1998). This choice is motivated by the model's particular emphasis on cognition and metacognition, as well as its recognition of the paramount importance of feedback, distinguishing it from other self-regulated learning (SRL) models.

Self-regulated learning (SRL), proposed by Dinsmore et al. (2008), comprises cognition, metacognition, and motivation. Cognition involves strategies for encoding, memorization, and recall. Metacognition enables understanding and monitoring of cognitive processes, while motivation influences the use and development of these skills. Butler & Winne (1995) emphasize the necessity of these components for enhancing SRL. For instance, when memorizing a task, metacognition helps learners select methods, monitor progress, and adjust strategies. Cognition involves employing learning strategies like repetition, and motivation determines the effort invested. These elements collectively contribute significantly to learning and are pivotal for self-regulated learning and academic success.

## 2.2 Reflective practices in learning

Enhancing self-regulated learning skills often involves promoting critical thinking through reflection, an integral aspect of the learning process. Reflection entails deliberate contemplation of past experiences to assess performance and gain fresh insights for future actions (Boud, Keogh, & Walker, 2013). Reflective writing, a common method, aims to teach self-reflection and takes various forms such as journals, online entries, essays, diaries, and portfolios (Barney & Mackinlay, 2010; McGuire, Lay, & Peters, 2009; Bruno & Dell'Aversana, 2017; Lo, 2010; Scott, 2010). Studies have demonstrated that reflective writing enables students to analyze their actions, assess performance, identify strengths and weaknesses, and recognize their developmental needs (Bjerkvik & Hilli, 2019). For instance, research by O'Loughlin and Griffith (2020) on reflective blogs in an anatomy course highlighted improvements in students' metacognitive skills, confidence, and professionalism. Similarly, Ramadhanti et al. (2020) found that reflective journals prompted metacognitive growth by encouraging students to respond to queries that raised awareness, enhanced self-evaluation, and supported monitoring of learning progress. These studies underscore the value of reflective writing in fostering independent learning strategies and improving metacognitive skills among students.

To summarize, Self-reflective writing can significantly enhance reflective practice and boost students' metacognitive awareness and motivation for learning. Although prior studies recognize the benefits of regular self-reflection assignments for improving Self-Regulated Learning (SRL), they haven't delved into the long-term effects of this practice. Therefore, this doctoral study aims to fill this gap by specifically examining how self-reflective writing influence students' SRL skills over time.

## 3 RESEARCH GOALS AND QUESTIONS

This doctoral research aims to investigate how frequent reflective writing impacts students' SRL (e.g. metacognitive regulation and judgement of learning, motivation and emotion). The main research questions addressed in this research are:

1) **RQ1:** How does frequent reflective writing affect metacognitive regulation and judgement of learning?
2) **RQ2:** How does frequent reflective writing affect emotion and motivation?

## 4 METHODOLOGY

The main theoretical framework for this research is Self-Regulated Learning Theory (SRL) that describes the process of independent learning. To facilitate student reflection, a framework derived from the D-I-E-P model (Cook, 1989) will be applied. This model, found beneficial for developing academic writing skills (Ono, A., & Ichii, R., 2019), includes four steps: Describe, Interpret, Evaluate, and Plan (depicted in Figure 2). These steps form the basis of reflective prompts in this study, aiming to engage students in the reflective process.

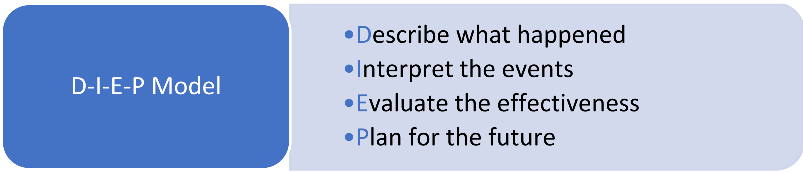



Figure 2: D-I-E-P Model adopted from Cook P. F. (1989)

The research, utilizing a case studies approach, centers on two distinct undergraduate courses at UniSA, Australia: 3rd-year engineering and 1st-year early childhood education. These courses were chosen due to their focus on cultivating self-regulated learning (SRL) skills. The third-year engineering students sought to enhance metacognitive awareness for their future careers, while early childhood education students aimed to develop effective self-reflection and learning judgment for teaching placements and subsequent courses. This study, targeting these disparate yet comparable courses, intends to investigate how regular reflective writing tasks can facilitate SRL skill acquisition across varied educational settings. In the autumn semester of 2023, enrolled students in any of these two courses undertook reflective writing tasks integrated into their assessments. These tasks included specific prompt questions following the D-I-E-P model (Cook, 1989).

The engineering course involved three reflective writing exercises. In the latter two exercises, students reflected on their instructor's feedback and prior assessment grades. Conversely, early childhood education students completed two exercises, the second one involving reflection on received feedback and grades. Students in both courses evaluated their learning strategies used in assessments and proposed modifications for future tasks. Furthermore, they estimated their expected grades, later compared with their actual marks to assess improvements in their judgment of learning through the reflective writing process over the semester. This study aims to discern the impact of reflective writing on students' SRL development within these distinct educational contexts.

Following Human Ethics Research approval, anonymized reflective writing exercises and assessment marks were obtained and analyzed using text analytics methods. The investigation aims to address research questions by conducting various analyses to scrutinize changes in each aspect of Self-Regulated Learning (SRL). For assessing metacognitive awareness, text analysis and semi-automated techniques guided by a large language model (LLM) will identify shifts in students' learning strategies within reflective exercises. Strategies will be categorized within each assessment, followed by an evaluation of their evolution and application correlated with students' performance and grades.

The assessment of students' judgment of learning (JOL) involves self-assessment and grade prediction against actual grades received after each assessment. The comparison between anticipated and actual grades seeks to identify significant differences, offering insights into potential improvements in JOL over time. Emotional analysis entails students' reflections on feelings regarding grades, assessment performance, and feedback. Large Language model (LLM) will analyze sentiment changes across the semester to detect emotional fluctuations influenced by reflective practices. Regarding self-efficacy, students' predictions of their final semester grades after reflective exercises will be scrutinized. This analysis aims to determine if regular reflection contribute to enhancing the alignment between students' self-efficacy beliefs and their actual academic performance.

## 5 CURRENT PROGRESS

Data collection is completed and I am in the process of analysing the data using an automated text analysis tool by language large model (LLM). Unique prompts have been precisely crafted for each research question, customized to the specific aspects of text analysis aligned with the research objectives. These prompts serve as the guiding principles for extracting valuable insights from the data and have undergone rigorous testing and refinement to ensure their effectiveness.

The testing phase involved applying the prompts to various examples of students' written reflections, thereby fine-tuning the approach. Moving forward, a dedicated Python program is in development for the analysis of extensive datasets of student responses. This program will enable efficient data processing and the generation of comprehensive results. The analytical process for each prompt encompasses multiple iterative steps, all aimed at ensuring the quality and reliability of the outcomes.

## 6 CONTRIBUTION

The primary motivation behind this doctoral project is to investigate how reflective writing exercises can improve students' SRL, namely their metacognition, learning startegies, self-efficacy, and judgement of learning. While previous research has explored enhancing SRK skills through self-reflective writing, an opportunity exists to systematically analyze reflective writing using automated text analytics tools. While numerous studies have indicated that consistent self-reflection and feedback on assignments can be instrumental in improving SRL, they have not extensively investigated the impact of sustained practice on subsequent reflective practices and the long-term enhancement of SRL. The completion of this project will provide a comprehensive understanding of the effects of these practices on students' SRL skills and contribute novel methodologies for analysing large amounts of written reflections using text analytics techniques.

## REFERENCES


Barney, K., & Mackinlay, E. (2010). Creating rainbows from words and transforming understandings: Enhancing student learning through reflective writing in an Aboriginal music course. *Teaching in Higher Education, 15*(2), 161-173. doi:10.1080/13562511003619995

Bjerkvik, L. K., & Hilli, Y. (2019). Reflective writing in undergraduate clinical nursing education: A literature review. *Nurse Education in Practice, 35*, 32-41. doi:10.1016/j.nepr.2018.11.013

Boekaerts, M. (1996). Self-regulated learning at the junction of cognition and motivation. European Psychologist. *European Psychologist, 1*(2), 100-112. doi:10.1027/1016-9040.1.2.100

Boud, D. (2010). *Assessment 2020: Seven Propositions for Assessment Reform in Higher Education.* Sydney: Australian Learning and Teaching Council.

Boud, D., Keogh, R., & Walker, D. (2013). Promoting reflection in learning: A model. In *Reflection: Turning Experience into Learning* (pp. 18-40). Routledge.

Bruno, A., & Dell'Aversana, G. (2017). Reflective Practice for Psychology Students: The Use of Reflective Journal Feedback in Higher Education. *Psychology Learning & Teaching, 16*(2), 248-260. doi:10.1177/1475725716686288

Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research, 65*(3), 245-281. doi:10.2307/1170684

Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research, 65*(3), 245-281. doi:10.2307/1170684

Chen, W., & Cheng, H. N. (2020). Analyzing Graduate Students' Behaviors of Self-regulated Learning in a Blended Learning Environment. *ACM International Conference Proceeding Series*, 53–57. doi:10.1145/3395245.3396416

Cheng, G. (2017). The impact of online automated feedback on students' reflective journal writing in an EFL course. *The Internet and Higher Education, 34*, 18-27. doi:10.1016/j.iheduc.2017.04.002

Cook, P. F. (1989). The Effect of Structured Training vs Less Formal Journal Writing on Quality of Thinking, Classroom Teaching Performance, and Attitudes toward Reflective Teaching during Preservice Teacher Training. *Annual Conference of the Northern Rocky Mountain Educational Research Association.* Jackson, Wyomin.

Delors, J. (1996). *Learning: the treasure within; report to UNESCO of the International Commission on Education for the Twenty-first Century.*

Dinsmore, D., Alexander, P., & Loughlin, S. (2008). Focusing the Conceptual Lens on Metacognition, Self-regulation, and Self-regulated Learning. *Educational Psychology Review, 20*(4), 391-409. doi:10.1007/s10648-008-9083-6

Efklides, A. (2011). Interactions of Metacognition With Motivation and Affect in Self-Regulated Learning: The MASRL Model. *Educational psychologist, 46*(1), 6-25. doi:10.1080/00461520.2011.538645

Jung, Y., & Wise, A. F. (2020). How and how well do students reflect?: multi-dimensional automated reflection assessment in health professions education. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, (pp. 595–604). Frankfurt Germany. doi:10.1145/3375462.3375528

Lew, D., & Schmidt, H. G. (2011). Writing to learn: Can reflection journals be used to promote self-reflection and learning? *Higher Education Research & Development, 30*(4), 519-532. doi:10.1080/07294360.2010.512627

Lo, Y.-F. (2010). Implementing reflective portfolios for promoting autonomous learning among EFL college students in Taiwan. *Language Teaching Research, 14*(1), 77-95. doi:0.1177/1362168809346509

McGuire, L., Lay, K., & Peters, J. (2009). Pedagogy of Reflective Writing in Professional Education. *Journal of the Scholarship of Teaching and Learning, 9*(1), 93-107.

O'Loughlin, V. D. (2020). Developing Student Metacognition through Reflective Writing in an Upper Level Undergraduate Anatomy Course. *Anatomical Sciences Education, 13*(6), 680-693. doi:10.1002/ase.1945

Ono, A., & Ichii, R. (2019). Business students' reflection on reflective writing assessments. *Journal of International Education in Business, 12*(2), 247-260.

Panadero, E., Broadbent, J., Boud, D., & Lodge, J. M. (2018). Using formative assessment to influence self- and co-regulated learning: the role of evaluative judgement. *European Journal of Psychology of Education, 34*(3), 535-557.

Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). Assessing Metacognition and Self-Regulated Learning. In P. R. Pintrich, C. A. Wolters, & G. P. Baxter, *Issues in the Measurement of Metacognition* (pp. 43–97). University of Nebraska.

Ramadhanti, D., Ghazali, A., Hasanah, M., Harsiati, T., & Yanda, D. P. (2020). The Use of Reflective Journal as a Tool for Monitoring of Metacognition Growth in Writing. *International Journal of Emerging Technologies in Learning (IJET), 15*(11), 162-187. doi:10.3991/ijet.v15i11.11939

Scott, S. G. (2010). Enhancing reflection skills through learning portfolios: An empirical test. *Journal of Management Education, 34*(3), 430-457. doi:10.1177/1052562909351144

Storie, B. (2021). *Incorporating Reflective Practice as a Means of Improving Student Self-Regulated Learning in a Digital Learning Environment.*

Vosniadou, S. (2020). ridging Secondary and Higher Education. The Importance of Self-regulated Learning. *The Importance of Self-Regulated Learning. European Review, 28(S1), S94-S103., 28*(S1), S94-S103. doi:10.1017/S1062798720000939

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* (Vol. 93, pp. 277–304). Routledge.

Zimmerman, B. J. (1990). Self-Regulated Learning and Academic Achievement: An Overview. *The Educational Psychologist, 25*(1), 3-17. doi:10.1207/s15326985ep2501_2

Zimmerman, B. J. (2002). Becoming a Self-Regulated Learner: An Overview. *Theory into practice, 41*(2), 64-70. doi:10.1207/s15430421tip4102_2

Zimmerman, B. J. (2008). Investigating Self-Regulation and Motivation: Historical Background, Methodological Developments, and Future Prospects. *American Educational Research Journal, 45*(1), 166-183. doi:10.3102/0002831207312909

# How might learning analytics be useful to identify and support curiosity in learning?

**Joe Tang**
Queensland University of Technology, Australia
q2.tang@qut.edu.au

**ABSTRACT**: This research addresses the intersection of curiosity and learning analytics. While curiosity enhances learning, there is a lack of data-driven studies on this topic. The study aims to explore and understand curiosity in learning, identify it using learning analytics, and support learners in becoming more curious. The research combines data-driven methods with behavioural trace data and self-reported information, offering real-time insights into learner behaviour. It involves four planned experiments, each focusing on specific aspects of curiosity and its impact on the learning process. This study aims to bridge the existing gap in our understanding of how curiosity can be harnessed to enhance the educational experience.

**Keywords**: Curiosity, Task-oriented learning, Behavioural trace data, Self-reported data

## 1 INTRODUCTION

Curiosity is widely acknowledged by educators and researchers as an enhancer of learning; accordingly, various disciplines such as psychology and education have long been investigating approaches and strategies to stimulate and sustain learner curiosity (Wu et al., 2018). Empirical studies (Kang et al., 2009; Lee et al., 2022) have found that curiosity has a positive impact on learners' learning experiences, including motivation to explore and remember new information, and persistence in learning. These qualities are also crucial for self-regulated learners who actively research what and how to learn to achieve their learning goals (Winne, 2017).

In terms of its impact on the study of curiosity, learning analytics can add new dimensions to our understanding of how learners approach new information and how their curiosity can be nurtured and supported. While the definition and measurement of curiosity remain subjects of ongoing debate in multi-disciplinary research (Hassin & Shohamy, 2020), the emerging field of learning analytics offers exciting possibilities for enhancing the learning experience and gaining new insights into learner engagement and needs (Viberg et al., 2018). However, despite the potential benefits, to the best of my knowledge, there have been no data-driven analytical attempts on studying curiosity.

In my view, the use of data-driven insights in learning analytics has the potential to bridge the gap between the research on curiosity and current educational practices (and beyond, such as lifelong learning). By leveraging learning analytics to support learner curiosity, new and innovative approaches can be developed to support learner curiosity. As part of my PhD project, I aim to fill the current gap in research on curiosity within the learning analytics field, both conceptually and analytically, focusing specifically on how learning analytics can be used to identify and support curiosity in learning.

## 2 RESEARCH GOALS AND QUESTIONS

The overarching goal of this research is to investigate the role of learning analytics in fostering greater curiosity among learners (related to RQ3). This is a critical question to be addressed in the learning analytics field. To address it effectively, a comprehensive understanding of the fundamental nature of curiosity in the context of learning is essential (related to RQ1). The focus of this study, which also presents a significant challenge, lies in establishing a connection between higher-level concepts like curiosity and the lower-level data available through analytics (related to RQ2). Overall, this doctoral project aims to answer the following research questions:

- **RQ1**: What is the nature of curiosity in learning?
- **RQ2**: How might learning analytics be useful to identify curiosity in learning?
- **RQ3**: How might learning analytics be used to support learners in becoming more curious?

## 3 CURRENT KNOWLEDGE AND RESEARCH GAPS

This doctoral project seeks to address gaps in both curiosity-driven learning and learning analytics: (1) it explores the underrepresented relationship between learner curiosity and learning within the field of learning analytics; (2) it introduces a data-driven approach to comprehensively investigate the multifaceted nature of curiosity in learning, leveraging the potential of learning analytics to bridge the existing research-practice gap.

### 3.1 Curiosity in Learning

Curiosity is a complex, multi-dimensional construct that can be understood in various ways (Grossnickle, 2016). Fundamentally, curiosity is the desire for new information or experiences in order to fill a knowledge gap or explore the unknown, and is often accompanied by emotion, increased arousal, or exploratory information-seeking behaviour (Grossnickle, 2016; Litman, 2005).

In my study, I adopt a concise definition from Wu et al. (2018, p. 920), who described *curiosity* as "*an emotionally induced, exploratory desire to solve a knowledge gap*". This definition succinctly highlights three crucial aspects of curiosity: (1) the existence of knowledge gaps is the main drive for curiosity (Loewenstein, 1994), (2) curiosity motivates exploratory information-seeking behaviour in novel, uncertain, or surprising environments (Ainley, 2019), and (3) curiosity involves different emotions during exploration (Litman, 2005).

Current research on connecting between curiosity and learning involves two main approaches, using self-reported assessments, such as questionnaires and surveys, and observing learner behaviours (Jirout & Klahr, 2012). While self-reported methods have shown some success in measuring curiosity, there are limitations and challenges in establishing the constructive validity of this data (Hadwin et al., 2007). Consequently, in recent decades, behavioural measures have gained popularity as a preferred approach for studying learning behaviours.

### 3.2 A Data-Driven Approach

With the increased demand for online learning modes, learners' interactions with their learning platforms are captured as log data and analysed to reveal patterns of learning behaviour, which can inform education practices (Gašević et al., 2015). Trace data have gained importance in learning analytics as they offer valuable insights into the learning process, particularly in the topic of self-regulated learning (SRL)(Winne, 2020). Recent studies indicated that trace data are more accurate in reflecting learner behaviours than self-reported data since the data are real-time and difficult for learners to alter (Ye & Pennisi, 2022). There is a growing trend of integrating behavioural trace data with theoretical frameworks to enhance and complement survey-based research in learning analytics studies. However, it is essential to note that this integration, as observed in recent research by Choi et al. (2023), posed a significant challenge as there was a misalignment between behavioural trace data and self-reported data.

Adopting existing and exploring new methods of utilising behavioural trace data, complemented with other types of self-reported data can be useful for studying curiosity. SRL is a particular topic in learning analytics that shares some common features with curiosity. In general, SRL encourages learners to actively research what they are learning and how to achieve their goals (Winne, 2017). Additionally, learners' self-regulation efforts towards their learning goals are influenced by multiple motivational factors such as interest, autonomy, proficiency and self-awareness (Schunk & Zimmerman, 2012). Even though it was not specifically mentioned, curiosity may also play a role towards SRL, as it has many overlapping relations with interest (Ainley, 2019; Grossnickle, 2016), and self-awareness (Goupil & Proust, 2023), among others.

### 3.3 Challenges

Building upon the discussion of adopting existing and exploring new methods for utilising behavioural trace data, as well as considering the potential role of curiosity in SRL, several challenges have emerged that need to be addressed:

(1) There is a misalignment between behavioural trace data and self-reported data. How can we adapt them to better understand learning and curiosity while avoiding potential biases?

(2) How can we bridge higher-level concepts, such as curiosity, with lower-level observations derived from trace and self-reported data? How can we ensure the learner behaviours we infer from trace data are really those behaviours?

## 4 METHODOLOGY

My research utilises a mixed-method approach, with the majority being quantitative so far. Data have been collected since July 2022 from two courses at the Queensland University of Technology (QUT) in Australia. Both are introductory courses on data analytics for undergraduate and postgraduate students, respectively. The courses teach students how to problem solve with data to extract business insight and make strategic decisions using Python programming language. JupyterLab[1] is utilised as the primary learning platform. Course materials and assignments are all accessible within this platform. Jupyter notebooks, a key component in JupyterLab, allow students to perform data analytics tasks and create notes, all directly within their web browsers.

Behavioural trace data have been collected from the Jupyter environment. By adopting the JupyterLab telemetry plugin[2] developed by the Educational Technology Collective group from the University of Michigan, the telemetry traces (14 types of events) are collected from the students when they interact with Jupyter notebooks. Each event is recorded in a JSON file with abundant metadata such as the names and types of notebooks.

The learning impact is evaluated by testing it during actual implementation in teaching activities. The task-oriented learning analytics model (Knight et al., 2020) is adopted as the framework, which takes into consideration the technical infrastructure and the learning impact in a task-centred learning environment. Consequently, students' self-reported data including learning confidence score data from learning tasks are currently collected via the telemetry plugin. Additionally, as part of their assessments, students are required to regularly document their reflections on their learning in the courses. These data will be used to validate the learning impacts and explore the emotional aspects of curiosity.

The research comprises a series of iterative experiments closely linked to the availability of actual classes and planned in accordance with the teaching schedules at QUT. Four experiments have been designed, as summarised in Table 1. Please note that, due to the nature of the iterative approach, RQ2 and RQ3 will be addressed through the processes of the experiments.

The university ethics committee has approved opt-out consent for the first 3 experiments, which means students are automatically included unless they choose to opt out. The process includes providing students with project information, outlining potential risks, and offering the option to opt out at any time. At the beginning of each semester, a recorded video presentation and a participant information sheet on the project are presented to the students. In the final week of the semester, initial findings from their data are presented.

## 5 CURRENT PROGRESS

I am currently in the second year of my PhD, having completed the literature review (to address RQ1) and successfully passed the confirmation of PhD candidature in March 2023. In addition, as summarised in Table 1, two experiments have been completed, with the third experiment currently in progress. I have presented two posters at the Australian Learning Analytics Summer Institute (ALASI) in 2022 and 2023, respectively.

[1] https://jupyter.org

[2] https://github.com/educational-technology-collective/etc_jupyterlab_telemetry_library

**Table 1: Summary of research experiment phases and current progress.**

| Experiment Phase | Implementation Period | Description | Status and Outcome |
|---|---|---|---|
| Pilot analysis on behavioural trace data | July 2022 – November 2022 | • In-depth exploration of data collection and cleaning processes.<br>• Understanding the structure and characteristics of trace data.<br>• Searching for and experimenting with exploratory data analysis techniques.<br>• Identifying potential traces of curiosity in the data. | • Completed<br>• Initial findings presented as a poster at ALASI 2022 |
| Implementing the task-oriented approach on the trace of exploratory behaviour | February 2023 – June 2023 | • Implement a task-centric approach with insight from the pilot analysis.<br>• Focusing on the interpretation of exploratory information-seeking behaviour.<br>• Deriving teaching strategies and interventions from the findings. | • Completed<br>• Journal paper in progress<br>• Poster presented at ALASI 2023 |
| Validating learning impact with self-reported data | July 2023 – November 2023; February 2024 – June 2024 | • Repeating the previous experiment.<br>• Designing and recording self-reported data from students regarding their confidence levels in completing learning content and tasks. | • Ethics approved<br>• Data collection in progress |
| Incorporating self-reflection on studying emotions & Closing the loop | February 2024 – June 2024; July 2024 – November 2024 | • Planned recruitment of student participants for an opt-in experiment.<br>• Integrating trace data, self-reported data (such as confidence levels and immediate self-reflection text entries), and ongoing self-reflection journals. | • Planned |

## REFERENCES


Ainley, M. (2019). Curiosity and interest: Emergence and divergence. *Educational Psychology Review*, *31*(4), 789–806. https://doi.org/10.1007/s10648-019-09495-z

Choi, H., Winne, P. H., Brooks, C., Li, W., & Shedden, K. (2023). Logs or Self-Reports? Misalignment Between Behavioral Trace Data and Surveys When Modeling Learner Achievement Goal Orientation. *LAK23: 13th International Learning Analytics and Knowledge Conference*, 11–21. https://doi.org/10.1145/3576050.3576052

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

Goupil, L., & Proust, J. (2023). Curiosity as a metacognitive feeling. *Cognition*, *231*, 105325. https://doi.org/10.1016/j.cognition.2022.105325

Grossnickle, E. M. (2016). Disentangling curiosity: Dimensionality, definitions, and distinctions from interest in educational contexts. *Educational Psychology Review*, *28*(1), 23–60. https://doi.org/10.1007/s10648-014-9294-y

Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, *2*(2), 107–124. https://doi.org/10.1007/s11409-007-9016-7

Hassin, R., & Shohamy, D. (2020). Editorial overview: Curiosity: Explore versus exploit. *Current Opinion in Behavioral Sciences*, *35*, iii–iv. https://doi.org/10.1016/j.cobeha.2020.10.004

Jirout, J., & Klahr, D. (2012). Children's scientific curiosity: In search of an operational definition of an elusive concept. *Developmental Review*, *32*(2), 125–160.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science*, *20*(8), 963–973. https://doi.org/10.1111/j.1467-9280.2009.02402.x

Knight, S., Gibson, A., & Shibani, A. (2020). Implementing learning analytics for learning impact: Taking tools to task. *The Internet and Higher Education*, *45*, 100729. https://doi.org/10.1016/j.iheduc.2020.100729

Lee, S. W.-Y., Hsu, Y.-T., & Cheng, K.-H. (2022). Do curious students learn more science in an immersive virtual reality environment? Exploring the impact of advance organizers and epistemic curiosity. *Computers & Education*, *182*, 104456. https://doi.org/10.1016/j.compedu.2022.104456

Litman, J. A. (2005). Curiosity and the pleasures of learning: Wanting and liking new information. *Cognition and Emotion*, *19*(6), 793–814. https://doi.org/10.1080/02699930541000101

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin*, *116*(1), 75.

Schunk, D. H., & Zimmerman, B. J. (2012). Motivation an essential dimension of self-regulated learning. In *Motivation and self-regulated learning* (pp. 13–42). Routledge.

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). Learning analytics for self-regulated learning. *Handbook of Learning Analytics*, 241–249.

Winne, P. H. (2020). Construct and consequential validity for learning analytics based on trace data. *Computers in Human Behavior*, *112*, 106457. https://doi.org/10.1016/j.chb.2020.106457

Wu, P.-H., Kuo, C.-Y., Wu, H.-K., Jen, T.-H., & Hsu, Y.-S. (2018). Learning benefits of secondary school students' inquiry-related curiosity: A cross-grade comparison of the relationships among learning experiences, curiosity, engagement, and inquiry abilities. *Science Education*, *102*(5), 917–950. https://doi.org/10.1002/sce.21456

Ye, D., & Pennisi, S. (2022). Using trace data to enhance students' self-regulation: A learning analytics perspective. *The Internet and Higher Education*, *54*, 100855. https://doi.org/10.1016/j.iheduc.2022.100855

# Designing Safe, Reliable, and Trustworthy Human-Centred Learning Analytics Systems

**Author(s): Riordan Dervin Alfredo**
Institution: Monash University
Email: riordan.alfredo@monash.edu

**ABSTRACT**: The proliferating convergence between learning analytics (LA) and artificial intelligence (AI) systems holds promise to enhance teaching and learning. However, several challenges remain in harmonising AI with human needs and control to foster safety, reliability, and trustworthiness when being appropriated in an authentic learning scenario. To address these challenges, my doctoral research seeks to balance human control and computer automation that can empower teachers and learners in AI-powered LA systems through co-creation practices. This research will be conducted through mixed-method research in healthcare simulation and aims to deepen our understanding of designing analytics and AI in LA systems with educational stakeholders guided by human-centred AI principles. My contributions will constitute (i) a systematic review of existing human-centred design practices and challenges in LA systems, (ii) human-centred design methods to co-create AI-powered systems guided by human-centred AI principles, and (iii) AI-powered LA systems appropriated by end-users with safety, reliability, and trustworthiness considerations. Ultimately, my research will establish human-centred design and deployment methodologies for AI-powered human-centred LA systems co-created with educational stakeholders in authentic educational settings.

**Keywords**: human-centred learning analytics; human-centred design; human-centred artificial intelligence; LA systems;

## 1 Introduction and Background

One of the important aspects of learning analytics is the creation of systems to support learning and teaching practices. These systems include data visualisation, LA dashboards, and other data-intensive systems (i.e., intelligent/AI-powered tools). These end-user solutions provide useful and actionable insights to help make pedagogical decisions, promote collaboration, improve feedback assessments, and encourage reflection and growth for learners, educators, and other educational stakeholders (Khosravi et al., 2022). With the rapid advancement of Artificial Intelligence (AI) technologies, the influence of AI is proliferating in the LA field (Carvalho et al., 2022; Siemens, 2013). These AI-powered analytics have the potential to revolutionise how educational data is collected, analysed, and used, with great promise to improve teaching and learning practices (Salas-Pilco et al., 2022).

Yet, the integration of AI in LA systems still poses several challenges, including data privacy, user-friendliness, and human agency between end-users (i.e., teachers and students) and AI-powered systems. In recent years, human-centredness has gained significant attention in LA research, leading towards the inception of the human-centred LA subfield (Martinez-Maldonado, 2023; Shum et al., 2019). Here, human-centred design (HCD) methods, such as participatory and co-design, have been widely used to create LA systems in collaboration with stakeholders (Sarmiento & Wise, 2022). In this sense, adopting HCD strategies in the design of AI-powered LA systems could help explore the balance between technical aspects of AI and human factors, such as teacher-student agency (Hooshyar et al., 2023; Lawrence et al., 2023), user experience (Bingley et al., 2023) and decision-making processes (Holstein et al., 2019). Designing practical AI-powered systems is also complex and requires multidisciplinary effort, especially in finding the right balance when involving educational stakeholders (teachers, students, and experts) with diverse expertise, needs, and values (Chen & Zhu, 2019; Dollinger et al., 2019; Martinez-Maldonado, 2023).

In the past few years, human-centred AI (HCAI) has emerged as a research topic to investigate human factors design and values aligned with the inherent challenges and concerns of AI technology (Xu, 2019). One particular view of this alignment has been proposed by Shneiderman (2022) in the HCAI framework. This framework prioritises human values such as **safety**, **reliability**, and **trustworthiness** in designing AI-powered systems and challenging end-user agency (control) over complete automation (Ozmen Garibay et al., 2023; Usmani et al., 2023). These three values are particularly important to current research in LA to accomplish the goals of understanding and supporting learning and teaching in practice (Chen & Zhu, 2019). Yet there are some limitations in current LA systems, such as a right balance respecting privacy between teachers and students, transparency in data sharing, and ensuring the accuracy of data-intensive insights to foster end-users trust and confidence (Drachsler & Greller, 2016; Nazaretsky et al., 2022).

## 2 Research Goals and Questions

My doctoral research explores the integration of HCD, AI, and LA fields to develop effective and practical AI-powered LA systems to empower educational stakeholders in an authentic learning setting. Situated in healthcare simulations, my research aims to deepen our understanding of various design stages in designing analytics and AI in LA systems with educational stakeholders. While healthcare simulations serve as a specific context for investigation, the findings are intended to be applicable broadly. The goals of this research are threefold: (i) to provide a systematic review of current human-centred design practices in LA systems, (ii) to formulate design methodologies for LA designers developing innovative AI-powered human-centred LA systems collaboratively with educational stakeholders that are guided by HCAI principles, and (iii) to provide evaluated artefacts through appropriation practices of AI-powered human-centred LA systems by end-users in authentic educational settings (i.e., healthcare simulations). This research poses the following question: "***How can we collaboratively design and deploy AI-powered learning analytics systems with educational stakeholders guided by human-centred AI principles?***". Specifically,

**RQ1.** What are the **existing practices and challenges**, such as stakeholder involvement, in designing human-centred learning analytics systems?

**RQ2.** How can human-centred AI principles (balanced human control and automation towards safety, reliability, and trustworthiness) guide the **collaborative design** of innovative AI-powered human-centred LA systems with teachers and students?

**RQ3**. How can AI-powered learning analytics systems **be appropriated by end-users** (teachers and students) while considering safety, reliability, and trustworthiness?

## 3 Current Problems and Solutions

In the realm of human-centred learning analytics, there exists a pressing need for a better approach to ensuring representative participation, considering expertise and lived experiences in design, balancing stakeholder input with technological innovation, and navigating power dynamics and decision-making processes (Martinez-Maldonado, 2023). Sarmiento & Wise (2022) conducted an initial non-systematic review of co-design and participatory design practices. While these practices are increasingly recognised as essential in the development of AI-powered LA systems, there exists a notable gap in the literature regarding the systematic evaluation of current stakeholder involvement practices and the challenges faced in LA research. It remains unclear the optimal strategies to involve educational stakeholders for balancing power and needs in the design process, as highlighted by Lang & Davis (2023). Given the proliferation of AI-powered LA systems, it is timely to conduct a systematic review that can identify gaps and provide recommendations in stakeholder involvement practices within human-centred LA research. Human-centred AI framework and principles may support this investigation (Shneiderman, 2022).

Despite a growing interest in creating AI-powered LA systems, many LA systems still lean towards a technology-centric approach. It may overlook human values and the necessity of human oversight in computer automation. Several works have proposed HCD methodologies to address this challenge and explore this opportunity. For example, Chen & Zhu (2019) explored how *value-sensitive design* can be employed in LA design to balance diverse human values, and Ahn et al. (2021) proposed a methodology to *co-design with stakeholders* concerning privacy, transparency, and security of the data used in LA. Recent works have proposed several design methodologies for AI-powered LA systems. For example, Lawrence et al., (2023) proposed six *design methods* to understand how teachers conceptualise sharing control with an AI co-orchestration tool that considers teacher agency, trust and reliability. Although some recent LA studies have adopted a human-centred design stance, there is still limited research on establishing design practices to aid LA designers in collaboratively designing AI-powered systems with stakeholders, specifically students and teachers. This highlights the need to establish design practices (e.g., methods, strategies, or principles) in gaining insights into stakeholders' perspectives on the agency, aligning their teaching or learning challenges with LA systems, and fostering discussions about safety, reliability, and trustworthiness. Situated in healthcare simulations, iteratively designing and developing AI-powered LA systems with nursing teachers and students may support this investigation, ultimately producing novel human-centred LA design methods.

Moreover, despite a growing interest in developing these AI-powered LA systems, little research has provided human-centred designed artefacts (i.e., systems) for researchers to explore the appropriation practices (Dix, 2007) of such solutions by end-users (teachers and students) in authentic learning environments. The continuous improvement between stakeholder partnerships may influence design choices and how they appropriate such solutions in real-world practices (Ahn et al., 2019). This may modify the intended use despite the systems being collaboratively designed with stakeholders, hindering integration in real-world usage. For instance, Nazaretsky et al. (2022) highlighted a challenge to design AI-powered LA systems that allow minimal shifts in teacher practice while ensuring the agency and privacy of students and teachers. It would be timely to investigate such appropriation practices of AI-powered LA systems to empower end-users with safety, reliability, and trustworthiness considerations (Usmani et al., 2023). Using healthcare simulations as a context of investigation and guided by HCAI principles, this exploration can lead to an understanding of how nursing teachers and students appropriate deployed systems in practice, such as design-in-use by co-evolving designed systems and usage adaptation (Kim & Lim, 2023; Nelson et al., 2008). By considering appropriation practices, end-users can potentially assess the safety, reliability, and trustworthiness of the systems.

This research holds significance in upholding human-centredness and ethical principles when supporting teaching and learning activities with co-created AI-powered human-centred LA systems. It seeks to recommend practices and challenges to involve stakeholders in designing LA systems, provides design methodological contributions for LA designers, and emphasises the importance of safety, reliability, and trustworthiness in appropriating AI-powered LA systems. Utilising empirical evidence and employing an iterative approach with nursing educational stakeholders, my findings are envisaged to aid researchers and practitioners in current and future developments of AI-powered LA systems. This research will contribute to the LA community by providing (i) a review of human-centred LA systems through the HCAI lens, (ii) practical and evidence-based human-centred LA design methods to involve stakeholders in the co-creation process, and (iii) AI-powered LA artefacts that are appropriated by end-users (teachers and students) with ethical considerations.

## 4 METHODS

To answer RQ1, a Systematic Literature Review (SLR) will be conducted to provide the foundation for this research to answer subsequent research questions (*RQ2* and *RQ3*). Although HCLA is becoming a trending topic within LA (Lang & Davis, 2023), the concept of human-centredness in LA literature has yet to be systematically reviewed. It aims to explore state-of-the-art design practices and challenges of LA systems through the HCAI lens. The outcomes will identify current stakeholder involvement techniques and challenges, literature gaps in the human agency and automation of LA systems, and offer recommendations to design human-centred AI-powered learning analytics systems with stakeholders that are safe, reliable, and trustworthy.

To answer RQ2, a longitudinal study will be conducted with nursing students and teachers, researchers, software developers, and LA designers to co-create an AI-powered healthcare simulation LA system (e.g., data visualisations and multimodal LA system, see Yan et al., 2023) guided by HCAI principles. This research will follow a design-based research method and HCD techniques (e.g., focus group, observation, co-design and participatory design), offering a deeper understanding of phenomena, generating hypotheses, and providing context and insights to inform collaborative design practice. The first iteration aims to develop a systematic design method to involve students who have completed clinical collaborative simulations but may have limited technical expertise in the design process. The second iteration will propose design principles when involving teachers and students in the co-creation process. The outcomes are novel and practical design practices, such as a design method or principle, addressing safety, reliability, and trustworthiness principles in human-centred LA, which will be supported by empirical evidence.

To answer RQ3, informed by design practices in RQ1 and RQ2, AI-powered human-centred LA systems can be co-created (iteratively developed and tested) with educational stakeholders as end-users (e.g., teachers and students) in authentic educational settings. Specifically, this research would be deployed in healthcare simulation scenarios to facilitate teamwork and communication learning. The appropriation of these deployed systems is further evaluated through mixed research methods (e.g., observation, interviews, and surveys). In the initial investigation, we seek to understand how nursing teachers would consider the safety, reliability, and trustworthiness of a stress visualisation dashboard on its usage in a stress-inducing simulation. Then, we seek to understand how nursing teachers transform features in a debrief-assistance multimodal LA system and adjust their teaching practices while using them in debriefing students after completing emergency clinical simulations. The insights of this co-evolution process can be explained using existing theoretical frameworks, such as *instrumental genesis* (Carvalho et al., 2019) and *design-in-use* (Kim & Lim, 2023). Ultimately, this research output can provide insights into how end-users design and use systems in real-world teaching or learning situations.

## 5 Current Status

Regarding ***RQ1***, I have written an article about a systematic review of human-centred learning analytics and AI in education, which is currently under internal review. Preliminary results indicate some consideration for human agency in LA systems design but limited end-user involvement in actual design and a lack of safety, reliability, and trustworthiness discussion when evaluating such systems. Regarding ***RQ2***, I have been adopting HCD methods to facilitate the co-design and deployment of AI-powered multimodal LA systems aligned with HCAI principles in the context of healthcare education. I have submitted a methodological paper to LAK24 aimed at helping LA designers and researchers align student challenges with AI-powered LA systems, understand student perspectives on agency, and facilitate discussions on safety, reliability, and trustworthiness. Future work will propose design guidelines to balance shared decision-making power between students, teachers, and researchers based on agency and pedagogical needs. Regarding ***RQ3***, I have authored an article at LAK23 (Alfredo et al., 2023) that made an initial exploration of nursing teachers' perceptions of trust towards an innovative AI-powered LA data visualisation. Currently, I am conducting iterative design studies to investigate how nursing teachers appropriate an AI-powered LA system in their practice to support students' reflection on a collaborative clinical simulation through an *instrumental genesis* lens. Future work will investigate how students appropriate an AI-powered LA dashboard to support self-reflection through the *design-in-use* lens.

## 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). https://doi.org/10.18608/jla.2019.62.5

Ahn, J., Campos, F., Nguyen, H., Hays, M., & Morrison, J. (2021). Co-Designing for Privacy, Transparency, and Trust in K-12 Learning Analytics. *LAK21: 11th International Learning Analytics and Knowledge Conference*, 55–65. https://doi.org/10.1145/3448139.3448145

Bingley, W. J., Curtis, C., Lockey, S., Bialkowski, A., Gillespie, N., Haslam, S. A., Ko, R. K. L., Steffens, N., Wiles, J., & Worthy, P. (2023). Where is the human in human-centered AI? Insights from developer priorities and user experiences. *Computers in Human Behavior*, *141*, 107617. https://doi.org/10.1016/j.chb.2022.107617

Carvalho, L., Martinez-Maldonado, R., & Goodyear, P. (2019). Instrumental genesis in the design studio. *International Journal of Computer-Supported Collaborative Learning*, *14*(1), 77–107. https://doi.org/10.1007/s11412-019-09294-2

Carvalho, L., Martinez-Maldonado, R., Tsai, Y.-S., Markauskaite, L., & De Laat, M. (2022). How can we design for learning in an AI world? *Computers and Education: Artificial Intelligence*, *3*, 100053. https://doi.org/10.1016/j.caeai.2022.100053

Chen, B., & Zhu, H. (2019). *Towards Value-Sensitive Learning Analytics Design* (WOS:000473277300046). 343–352. https://doi.org/10.1145/3303772.3303798

Dix, A. (2007). *Designing for Appropriation*. Proceedings of HCI 2007 The 21st British HCI Group Annual Conference University of Lancaster, UK (HCI). https://doi.org/10.14236/ewic/HCI2007.53

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). https://doi.org/10.18608/jla.2019.62.2

Drachsler, H., & Greller, W. (2016). Privacy and analytics: It's a DELICATE issue a checklist for trusted learning analytics. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 89–98. https://doi.org/10.1145/2883851.2883893

Holstein, K., McLaren, B. M., & Aleven, V. (2019). Co-Designing a Real-Time Classroom Orchestration Tool to Support Teacher–AI Complementarity. *Journal of Learning Analytics*, *6*(2). https://doi.org/10.18608/jla.2019.62.3

Hooshyar, D., Tammets, K., Ley, T., Aus, K., & Kollom, K. (2023). Learning Analytics in Supporting Student Agency: A Systematic Review. *Sustainability*, *15*(18), Article 18. https://doi.org/10.3390/su151813662

Kim, H., & Lim, Y.-K. (2023). Investigating How Users Design Everyday Intelligent Systems in Use. *Proceedings of the 2023 ACM Designing Interactive Systems Conference*, 702–711. https://doi.org/10.1145/3563657.3596039

Lang, C., & Davis, L. (2023). Learning Analytics and Stakeholder Inclusion: What do We Mean When We Say "Human-Centered"? *LAK23: 13th International Learning Analytics and Knowledge Conference*, 411–417. https://doi.org/10.1145/3576050.3576110

Lawrence, L., Echeverria, V., Yang, K., Aleven, V., & Rummel, N. (2023). How teachers conceptualise shared control with an AI co-orchestration tool: A multiyear teacher-centred design process. *British Journal of Educational Technology*, *n/a*(n/a). https://doi.org/10.1111/bjet.13372

Martinez-Maldonado, R. (2023). Human-centred learning analytics: Four challenges in realising the potential. *Learning Letters*, *1*, 6–6. https://doi.org/10.59453/FIZJ7007

Nazaretsky, T., Cukurova, M., & Alexandron, G. (2022). An Instrument for Measuring Teachers' Trust in AI-Based Educational Technology. *LAK22: 12th International Learning Analytics and Knowledge Conference*, 56–66. https://doi.org/10.1145/3506860.3506866

Nelson, J., Buisine, S., & Aoussat, A. (2008). *Design in use: Some methodological considerations*. https://api.semanticscholar.org/CorpusID:167215398

Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., Falco, G., Fiore, S. M., Garibay, I., Grieman, K., Havens, J. C., Jirotka, M., Kacorri, H., Karwowski, W., Kider, J., Konstan, J., Koon, S., Lopez-Gonzalez, M., Maifeld-Carucci, I., … Xu, W. (2023). Six Human-Centered Artificial Intelligence Grand Challenges. *International Journal of Human–Computer Interaction*, *39*(3), 391–437. https://doi.org/10.1080/10447318.2022.2153320

Salas-Pilco, S. Z., Xiao, K., & Hu, X. (2022). Artificial intelligence and learning analytics in teacher education: A systematic review. *Education Sciences*, *12*(8), 569.

Sarmiento, J. P., & Wise, A. F. (2022). Participatory and Co-Design of Learning Analytics: An Initial Review of the Literature. *LAK22: 12th International Learning Analytics and Knowledge Conference*, *1*, 535–541. https://doi.org/10.1145/3506860.3506910

Shneiderman, B. (2022). *Human-Centered AI*. Oxford University Press.

Shum, S. B., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-Centred Learning Analytics. *Journal of Learning Analytics*, *6*(2), Article 2. https://doi.org/10.18608/jla.2019.62.1

Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, *57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Usmani, U. A., Happonen, A., & Watada, J. (2023). Human-Centered Artificial Intelligence: Designing for User Empowerment and Ethical Considerations. *2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 01–05. https://doi.org/10.1109/HORA58378.2023.10156761

Xu, W. (2019). Toward human-centered AI: A perspective from human-computer interaction. *Interactions*, *26*(4), 42–46. https://doi.org/10.1145/3328485

Yan, L., Martinez-Maldonado, R., Zhao, L., Dix, S., Jaggard, H., Wotherspoon, R., Li, X., & Gašević, D. (2023). The role of indoor positioning analytics in assessment of simulation-based learning. *British Journal of Educational Technology*, *54*(1), 267–292. https://doi.org/10.1111/bjet.13262

# Scaffolding Feedback Literacy and Self-Regulated Learning in Learning Analytics-enabled Feedback Processes

**Hua Jin**
Monash University
Flora.jin@monash.edu

**ABSTRACT**: Feedback is essential for learning. The emerging concept of feedback literacy underscores the skills students require for effectively utilising feedback to regulate their learning. This highlights the importance of self-regulated learning (SRL) skills in the feedback process since SRL processes involve a learner's ability to control, monitor, and regulate their learning, enabling students to actively engage with and make use of feedback. Learning Analytics (LA) can offer insights into students' learning engagement by exploiting the evidence generated throughout the SRL processes while students interact with content and educational resources, which can potentially scaffold feedback literacy and SRL skills. However, there isn't a clear mapping between the specific dimensions of feedback literacy and the diverse processes of SRL, which limits the full potential of LA in scaffolding feedback literacy and SRL. Against this gap, this doctoral research aims to study the relationship between feedback literacy and SRL tightly with the goal of supporting them through the use of LA.

**Keywords**: Learning analytics, Self-regulated learning, Feedback process, Feedback literacy

## 1. INTRODUCTION

Feedback is an integral component of education, playing a crucial role in the teaching and learning process. However, various studies have reported that students often struggle to interpret and act on feedback and generally express dissatisfaction and a lack of engagement with the feedback process (O'Donovan et al., 2021). These challenges can be attributed to students' lack of feedback literacy, which refers to the skills, understandings, and mindsets that enable students to effectively make sense of and act on feedback, thereby facilitating the uptake of feedback (Carless & Boud, 2018). Nevertheless, students' ability to uptake feedback is seriously constrained unless students are capable of self-regulating their own learning, which is an essential aspect of making feedback sustainable (Carless et al., 2011). This underscores the role of self-regulated learning (SRL) in the feedback process, which is a concept that describes learners' cognitive, motivational, and emotional facets that enable learners to self direct their own learning process (Panadero, 2017). For this purpose, a seminal study (Butler & Winne, 1995) explored the relationship between SRL and feedback, recognising feedback's multifaceted roles in promoting SRL and suggesting that SRL and feedback should be tightly linked.

Learning Analytics (LA) has emerged as a promising solution to facilitate the feedback process and support SRL (Jin et al., 2022). However, current LA feedback tools tend to focus on delivering data-driven feedback on learners' engagement with learning activities or materials (such as page viewing and time on tasks), rather than focusing on understanding how students make sense of feedback and act on it (Winstone, 2019). This focus impedes the effective support for student feedback literacy. Additionally, a systematic review of LA tools revealed their limitations in supporting SRL and their lack of consideration for SRL theories (Matcha et al., 2019). These limitations in existing

LA feedback tools, coupled with a lack of consideration of learners' perspectives (Jivet et al., 2018) and educational theories in their design (Tsai, 2022), have arguably hindered a holistic understanding of the impact of feedback and its potential to scaffold feedback literacy and SRL skills.

In light of these challenges, a more granular approach to scaffolding feedback literacy is essential, incorporating essential SRL processes such as goal setting, monitoring, evaluation, and reflection. Given the interplay between feedback literacy and SRL, and their significance in enabling students to effectively utilise feedback, this doctoral research aims to deeply investigate the interplay between feedback literacy and SRL using LA. This will consider learners' perspectives and feedback theories, informing the design of an LA feedback tool to more effectively scaffold feedback literacy and SRL.

## 2. LITERATURE REVIEW

### 2.1 Feedback literacy and self-regulated learning

The importance of feedback cannot be stressed enough in learning. In recent years, studies on effective feedback tend to shift the concept of feedback from a traditional transmission-focused model to a reciprocal model (Boud & Molloy, 2013), recognising the importance of two-way communication process between educators and learners, enabling the feedback process more attuned to individual needs (Yang & Carless, 2013). In this dialogic approach, it is essential for students to take an active role. This participation allows them to deeply engage in the feedback dialogue, which in turn builds trust and motivates further engagement with feedback (Yang & Carless, 2013). Yet, existing literature consistently points out students' challenges in engagement with feedback, stemming from difficulties in interpreting and applying feedback to subsequent work, leading to a reliance on educators for more explicit and prescriptive guidance (O'Donovan et al., 2021). These challenges may be attributed to students' lack of feedback literacy, which encapsulates the capabilities, attitudes, dispositions, and mindsets students require to make effective use of feedback (Carless & Boud, 2018). For feedback to be effective, it is crucial that students possess a certain degree of feedback literacy for them to be proactive and self-directed in their engagement with and utilisation of feedback (Carless & Boud, 2018).

The concept of feedback literacy has been explored through multiple frameworks, each providing a distinct lens to understand the multifaceted nature of student engagement with feedback. Carless and Boud (2018) define feedback literacy as encompassing the understanding, capacities, and dispositions needed to interpret and implement feedback. Their framework includes four core features: *appreciating feedback*, *making judgments*, *managing affect,* and *taking action*. In more recent research, Molloy et al., (2020) expanded upon this concept through a large-scale survey and qualitative study, culminating in a more comprehensive framework for feedback literacy. They proposed a set of seven aspects, including (1) *committing to feedback as improvement*, (2) *appreciating feedback as an active process*, (3) *eliciting information to improve learning*, (4) *processing feedback information,* (5) *acknowledging and working with emotions*, (6) *acknowledging feedback as a reciprocal process*, and (7) *enacting outcomes of processing of feedback information*. In comparison with the four features proposed by Carless and Boud (2018), the framework by Molloy et al. (2020) adds depth by identifying the dispositions, beliefs, and approaches to feedback. This framework places particular emphasis on the importance of self-regulation skills, highlighting the connection between self-regulation skills and the development of aforementioned aspects.

To become feedback literate, self-regulated learning (SRL) skills are crucial and fundamental. For example, a study (Carless & Boud, 2018) pointed out that if students lack the capacities to self-regulate their own learning, their ability to make sense of and use any feedback that they receive is seriously constrained. This implies that SRL plays a crucial role in fostering student feedback literacy and impact (Molloy et al., 2020). Self-regulated students engage with academic tasks, set goals and attempt to achieve them by applying tactics and strategies under the self-monitoring process since SRL is a "deliberate, judgemental, adaptive process" (Butler & Winne, 1995) to "transform their mental abilities into academic skills" (Zimmerman, 2002). The intrinsic link between feedback literacy and SRL is evident when delving into the nuances of both concepts. For example, students' emotional resilience (*Managing affect*), as an affective dimension of self-regulation, is one of the capabilities that students need in order to properly deal with feedback; students' ability to evaluate (*Making judgment*) their own work and feedback information, which is related to the judgemental dimension of SRL, enabling them to self-evaluate their performance and decide which learning strategy or tactic to implement (*Taking action*). Therefore, SRL plays a crucial role in promoting uptake of feedback since it directly impacts the development of feedback literacy.

While numerous studies have explored components of student feedback literacy, how we can scaffold the development of these components remains unclear. This is partly due to the challenge to effectively track how students interact with feedback (Jin et al., 2022; Winstone, 2019), how they interpret, internalise, and act upon the feedback they receive, as well as how that feedback ultimately affects their learning. Moreover, existing studies on the development of feedback literacy focus on its integration within course design and the role of teachers in facilitating this process. However, there is a limited understanding at a granular level of how specific dimensions of feedback literacy correspond with specific SRL processes. This missing linkage between feedback literacy and SRL hampers the effective teaching support needed for the development of both student feedback literacy and SRL skills. In light of this, this doctoral research proposes to utilise LA to capture and analyse student interactions with feedback to prompt SRL processes (e.g., goal-setting, monitoring, evaluation, reflection) among students and scaffold their feedback literacy. Therefore, it is imperative to understand students' experiences with feedback, particularly their sense-making and action-taking processes, informing the design of an LA feedback management tool, that is grounded in theories of effective feedback. As such, the first research question (RQ1) is: **What can students' experiences with feedback reveal about their feedback literacy and self-regulated learning skills?**

## 2.2 Learning analytics, feedback and self-regulated learning

In recent years, Learning Analytics (LA) has gained prominence as a technology-enhanced approach to facilitating the feedback process (Yang & Carless, 2013) and tracking feedback impact (Winstone, 2019). It leverages extensive learning data to provide insights into student engagement in digital learning environments. By utilising diverse online engagement data and performance data available within the platforms (e.g., Learning management systems), LA tools aim to enhance students' overall experience with feedback focusing on improving feedback content, frequency, and timeliness. However, little attention is paid to the sense-making process of feedback or the support needed to drive students to turn feedback into action (SRL) (Winstone, 2019), thus leaving a gap in the understanding of feedback effectiveness, specifically the extent to which feedback is used by learners to self-regulate their learning. Additionally, a systematic review (Matcha et al., 2019) of 29 LA tools pointed out that no paper explicitly considered SRL theories in the design of the feedback tools, and the majority of them focused on online engagement such as number of logins, posts, or questions answered rather than targeting at self-regulation level. The issue of stakeholder engagement and buy-in have also been found in the study (Tsai et al., 2020) due to students' lack of involvement in the design and implementation of LA. Considering such challenges, a human-centred approach to the LA feedback tool is needed since it involves incorporating the user's perspective into the design process in order to fully understand their needs, as well as the educational theories that underpin the tools (Buckingham Shum et al., 2019).

To understand the interplay between feedback literacy and SRL processes through the use of an LA tool, we first investigate how students with varying levels of feedback literacy skills demonstrate SRL processes and engage with the LA based feedback process. Thus, the second research question (RQ2) is: **How can LA be utilised to identify the interplay between feedback literacy and SRL processes?**

This question will allow us to explore how we could use the LA feedback tool to support SRL processes, allowing students to make effective use of feedback and ultimately enhance their feedback literacy. As such, the third research question (RQ3) is: **How may an LA feedback management tool scaffold feedback literacy through the development of SRL skills?**

## 3. METHODOLOGY

### 3.1 Research question 1

Two surveys and two focus group/interview sessions were conducted to understand both educators' and students' current feedback practices. These sessions delved into their perspectives on effective feedback and data-driven feedback, along with their perceived challenges. A global survey with 282 educator respondents from 27 countries provided insights into their feedback practices. This was followed by in-depth interviews with 20 educators to gain a deeper understanding of their perceptions and challenges in feedback processes. On the student side, a survey gathered feedback experiences from 641 students, focusing on their challenges (sense-making and action-taking processes) and motivations in the feedback process. Subsequently, eight focus groups with 36 students were formed to discuss feedback practices, challenges, and their views on data-driven feedback. These sessions also involved brainstorming desired features for a feedback tool. Both the educator and student data was analysed using thematic analysis on NVivo software. To answer RQ1, all studies primarily used qualitative methods, such as surveys and interviews/focus groups. This enabled us to explore students' feedback literacy skills and challenges associated with tracking feedback impacts based on their experiences. This insight subsequently informed the design of an LA feedback management tool named PolyFeed. The tool contains three key functions: 'Annotate Feedback', 'Create work plan', and 'View summary'.

### 3.2 Research question 2

PolyFeed will be developed further to address RQ2. Specifically, PolyFeed collects data about students' sense-making of feedback (Annotate Feedback) and their formulation of actions (Create Work plan). Subsequently, the tool generates analytics regarding the common strengths, weaknesses, actions across courses, assignments, and peers, along with insights into individual learning behaviours (Feedback Analytic). All trace data (students' interactions with feedback) collected through the tool will help elucidate the relationship between feedback literacy and SRL processes. As part of the answer to research question 2, we will also conduct another field study, which is to cluster students' feedback literacy skills to examine how students with varying levels of feedback literacy skills demonstrate SRL processes on FLoRA platform (Facilitating Self-Regulated Learning with Personalised Scaffolds on Student's own Regulation Activities). In short, the platform is grounded in SRL theories and aims to scaffold students' SRL skills. It has been utilised for data collection across several countries for many years. Students are required to write a 300-400 word essay based on reading materials on the FLoRa platform. While completing the task, they are allowed to use tools provided by the platform, such as annotations, highlights, a time-planner, and a timer. Those trace data will be used to analyse their learning strategies and SRL skills. To understand the relationship between SRL and feedback literacy, a feedback literacy scale (work in progress) will be incorporated into FLoRa platform. This will enable observation of how students demonstrate different SRL processes across various levels of feedback literacy, as evidenced through trace data. Additionally, interviews will be conducted and will be targeted their SRL process to understand how LA-based feedback (ChatGPT feedback designed to target students' SRL processes and conditions) made impact on their SRL and feedback literacy skills, for example, changes in their emotional responses, evaluation process and cognitive process (COPE model). Insights from both studies will answer RQ2.

### 3.3 Research question 3

From the two studies proposed in RQ2, our goal is to establish a clear mapping between feedback literacy and SRL processes. This mapping will subsequently be incorporated into the LA feedback tool (PolyFeed) to enhance its functionalities, allowing it to effectively target various SRL processes, thereby scaffolding students' feedback literacy. This approach will address RQ3.

## 4. PRELIMINARY FINDINGS AND PUBLICATIONS

Studies involved in RQ1 explored the multifaceted dynamics of student feedback literacy by understanding the experiences, perceptions, and interactions of students with feedback. Our preliminary findings revealed a noticeable discrepancy in the feedback literacy skills demonstrated by students. They exhibited a significant degree of feedback literacy skills when discussing their current feedback practices in making use of feedback. However, their articulation of the challenges within the feedback process was predominantly attributed to challenges created by external factors (e.g., learning design, feedback design). Additionally, significant differences were observed in their inclination to act upon feedback and in the challenges experienced by students with different levels of feedback literacy skills. Educators' challenges and perceived effective elements in the feedback processes were also observed. Based on all above findings, an LA feedback management tool was designed and validated, which will be developed further and utilised to explore the interplay between feedback literacy and SRL processes.

All publications related to this doctoral research (first author) are summarised below:

• One full conference paper was published to ASCILITE 2022 (Australasian Society for Computers in Learning in Tertiary Education), and received the **best full student paper award**. **Title**: **Towards**

**Supporting Dialogic Feedback Processes Using Learning Analytics: the Educators' views on Effective Feedback.**

• One journal paper is under review (Instructional Science). **Title: Exploring Student Feedback Literacy: Students' Self-reported Experiences with Feedback Processes.**

• One full conference paper is submitted to the 14th International Conference on Learning Analytics and Knowledge (LAK'24). **Title:Scaffolding Feedback Literacy: Designing a Feedback Analytics Tool with Students.**

• One paper is in progress and is related to RQ1, which investigates educators' current practice with feedback based on survey data. **Title: Feedback in K-12 and Higher Education: Educators' Perceptions and Implications for Learning Analytics.**

## REFERENCES


Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in higher education*, *38*(6), 698– 712.

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

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research*, *65*(3), 245–281.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315– 1325.

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in higher education*, *36*(4), 395–407.

Jin, H., Martinez-Maldonado, R., Li, T., Chan, P. W. K., & Tsai, Y.-S. (2022). Towards supporting dialogic feedback processes using learning analytics: the educators' views on effective feedback. *ASCILITE Publications*, e22054–e22054.

Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018). License to evaluate: Preparing learning analytics dashboards for educational practice. In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 31–40).

Matcha, W., Gaˇsevi´c, D., Pardo, A., et al. (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.

Molloy, E., Boud, D., & Henderson, M. (2020). Developing a learning-centred framework for feedback literacy. *Assessment & Evaluation in Higher Education, 45*(4), 527–540.

O'Donovan, B. M., den Outer, B., Price, M., & Lloyd, A. (2021). What makes good feedback good? *Studies in Higher Education, 46*(2), 318–329.

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in psychology*, 422.

Tsai, Y.-S. (2022). Why feedback literacy matters for learning analytics. *arXiv preprint arXiv:2209.00879.*

Tsai, Y.-S., Rates, D., Moreno-Marcos, P. M., Mun˜oz-Merino, P. J., Jivet, I., Scheffel, M., . . . Gašević, D. (2020). Learning analytics in European higher education—trends and barriers. *Computers & Education, 155*, 103933.

Winstone, N. (2019). Facilitating students' use of feedback: Capturing and tracking impact using digital tools. In *The impact of feedback in higher education* (pp. 225–242). Springer.

Yang, M., & Carless, D. (2013). The feedback triangle and the enhancement of dialogic feedback processes. *Teaching in Higher Education, 18*(3), 285–297.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into practice, 41*(2), 64–70.

# Goal Setting and Academic Performance: Using a Conversational Agent to Support SRL Processes for Higher Education Students

**Gabrielle Martins van Jaarsveld**
Erasmus University Rotterdam
martinsvanjaarsveld@essb.eur.nl

**ABSTRACT**: The increasing prevalence of technology in higher education has highlighted the importance of self-regulated learning skills for students in modern educational settings. The presence of this technology has also increased the ease with which large amounts of data can be collected about student behavior and performance and used to offer them personalized support. In this project, we aim to explore how learning analytics can be leveraged to support self-regulated learning in higher education environments. Using a design-based approach, we have built a conversational agent to carryout self-regulated learning interventions. These interventions are built to scaffold the self-regulated learning cycle, with different forms of text-mining of student interactions with the conversational agent being used to inform the real-time metacognitive feedback offered to students about the quality of their self-regulated learning process. Initial findings from this project show that scaffolding and feedback during the primary phase of the self-regulated learning cycle improve the quality of the process, but alone, they do not improve learning outcomes. In the next phase of this project, we aim to extend the scaffolding and feedback implementation to later phases of the self-regulated learning cycle, to improve students self-regulated learning processes, and thus indirectly, students' academic performance.

**Keywords**: Self-regulated learning, goal setting, conversational agent, learning analytics, feedback

## 1 INTRODUCTION

The past decades have seen higher education undergoing a fast shift towards digitalization, with ICT becoming an increasingly common part of student's day-to-day study activities (Uğur & Guliz Ugur, 2020). This shift brought attention to the importance of self-regulated learning to succeed in learning environments that give students greater autonomy over when and how to learn with less teacher oversight. Studies have shown that SRL skills are strongly related to student performance within higher education (Zimmerman & Schunk, 2011), especially within digital learning environments (Broadbent & Poon, 2015; Wong, Baars, Davis, et al., 2019). The sudden shift to online education during the COVID-19 pandemic highlights the importance of SRL skills in educational environments which are increasingly incorporating digital tools. However, this increasing digital educational landscape also has its affordances: it allows for the creation and implementation of more scalable and personalized support tools, informed by the wide array of data that can efficiently be collected and analyzed.

The overarching research question of this project is: how can learning analytics be used to support self-regulated learning in higher education environments?

## 2 CURRENT KNOWLEDGE AND CONCEPTUAL FRAMEWORK

### 2.1 Goal Setting within the Self-Regulated Learning Framework

There are many theoretical approaches to the concept of SRL. However, Zimmerman's triarchic model of SRL is among one of the most common models used in educational research, and provides a base upon which several other models expand (Panadero, 2017). In this model, Zimmerman (2011) describes SRL as a cyclical process in which students move through three stages, 1) the forethought phase, in which they set goals and create a plan; 2) the performance phase, in which they carry out study activities, and monitor their progress; 3) and the reflection phase, in which they reflect on their progress and adapt it for future iterations of the cycle. Goal setting forms the base of this model, as well as many other common models of SRL, and drives all subsequent phases within the cycle (Panadero, 2017). As a result, goal-setting interventions are a common approach to supporting and improving students' SRL skills as they allow early intervention in the cycle to develop a strong basis upon which the subsequent SRL phases can build.

However, while there is a lot of literature in which goal-setting interventions are implemented in higher education, there is little consensus on how best to support this process. Furthermore, very little literature on the topic thoroughly evaluates the tools they use to ensure they have the intended outcomes (Martins van Jaarsveld et al., 2023a). Some studies have found that goal-setting interventions can be an effective method of improving academic performance. Still overall, the studies on this topic have mixed results and the outcomes seem to heavily depend on the context and content of the intervention itself (Latham & Locke, 2007; Martins van Jaarsveld et al., 2023a). Furthermore, while goal-setting interventions may form a good basis for supporting SRL, offering additional support for the later stages of the cycle, namely the performance and reflection phase, has also been suggested as an important element of effective SRL support tools (Martins van Jaarsveld et al., 2023a; Wong, Baars, Davis, et al., 2019). There is, therefore, a need for further research on how best to design and implement goal-setting interventions to support SRL processes, and subsequently, academic performance.

### 2.2 Conversational Agents

Studies that focused on digital delivery of educational interventions reflect the fact that digital interventions are often easier for participants to access and easier for the providers to scale to large populations of students (Khosrawi-Rad et al., 2022; Warriem et al., 2022). However, this is not without its weak points. Studies have shown that SRL interventions are more effective when carried out by a researcher than when participants use static, self-led versions (Wang et al., 2021). By creating a scalable digital solution, conversational agents may address this gap between researcher and participant-led interventions. Conversational agents, or chatbots, are tools which individuals can interact with in the form of a dialogue, allowing pre-structured interactions to take the form of a conversation, rather than a static form (McTear et al., 2016). Given the more natural state of interaction participants can have with these kinds of tools, they mimic researcher-led interventions more closely than a traditional static intervention, since they allow the tool to drive, structure, and prompt the interaction. Additionally, a common problem with SRL support tools is that of engagement, and how those students most in need of support are often the ones least likely to use the support tools (Ryan et al., 2001). Studies have shown that conversational agents can be more engaging and motivating than their static counterparts (Hew et al., 2022; Perski et al., 2019), and since they can be used to initiate and lead interventions, not relying on participants to drive the process, they can therefore be used to target populations who tend to have lower levels of engagement. Therefore, within this research project, we aim to use conversational agents to deliver SRL support interventions to make it more easily scalable and adaptable, while also increasing engagement and decreasing the barrier of access for at-risk students or students with lower SRL skills.

### 2.3 Learning Analytics

Learning analytics (LA) commonly refers to the collection and use of data about learners and their contexts to understand and support their learning processes (Tsai, 2019). More practically speaking, LA involves the collection of data from students' interactions with digital learning tools and materials, and using this to understand or support students. LA can be highly impactful in education, and there has been a lot of research specifically into how learning analytics can be used to support students SRL skills (Heikkinen et al., 2022). However, according to a systematic review by Wong et al. (2019), much of the existing research on this topic follows a top-down approach to designing learning analytics tools, with very few tools being built based on educational theory. Instead, this study shows a trend in LA research in which tools are designed based on what data is available, and not necessarily based on what supports students need or want, (Wong, Baars, Koning, et al., 2019). This is a common critique of LA, and the dashboards which often result from these data, as they generally focus on behavioral analytics since this is the data which is commonly collected and easily accessible, with little focus on deeper metacognitive analytics and a lack of support for the validity of these measures. While LA feedback at a cognitive and behavioral level might be useful in some situations, studies have suggested that feedback at a metacognitive level may be more beneficial, especially in supporting SRL and improving SRL processes (Lee et al., 2010). Based on the existing research on this topic and the current gaps in the literature, with this project, we aim to a) build LA informed feedback using a bottom-up approach based on educational theory, and b) use LA to offer personalized feedback to students at a metacognitive level, to support SRL processes, and thus indirectly improve academic performance.

## 3 SUGGESTED SOLUTION

As a solution to the problems outlined in the previous section, for this project, we created a conversational agent to carry out a goal-setting intervention, using learning analytics to provide personalized metacognitive level feedback to students throughout the process. We propose that the intervention and feedback be designed using a bottom-up approach based on educational theory and existing research. The goal-setting activity, SRL feedback, and all other guided SRL processes should be carried out via a conversational agent, allowing for a more engaging but dialogue-rich interaction. The goal-setting intervention, should be supported with feedback and additional elements in the intervention that follow through with the SRL cycle, providing support and feedback for the performance and reflection phases (Heikkinen et al., 2022). With this proposed solution, we aim to improve the quality of students' SRL process and, in turn, improve student performance.

## 4 RESEARCH QUESTIONS AND METHODOLOGY

This project will follow a design-based research approach, meaning that the development and implementation of the SRL support tool will be iterative, with each stage of the research building on the findings from the previous stage. The feedback offered during the interventions is called "SRL analytics", derived from students' descriptions of and reports on their SRL activities. This feedback will, therefore, be primarily derived from text mining, or text analytics, in which student dialogue with the conversational agent will be processed first manually via coders to create and validate coding schemes and then automatically, via machine learning modules to extract information about the quality and content of their SRL processes and give personalized feedback in response. This research consists of four studies, with each one building on the findings of the previous. All experimental studies will be carried out within student populations from the bachelor and masters' programs of a social sciences faculty of a large Dutch university. Participation in the experiments will be voluntary, and the experimental activities will be carried out outside the classroom in students own study environments, in their own time . A brief overview of each of these studies and the associated research questions can be found in the sections below.

### 4.1 Study 1: Overview and RQs

To better understand the current state of the literature on goal setting in higher education settings, Study 1 consists of a systematic literature review, examining papers published after 2009 in which academic goal-setting activities were carried out in higher education settings. The following research questions were examined in this paper:

1. What are the elements of goal-setting interventions that have been carried out in previous studies in higher education institutions?
2. What is the effect of goal-setting interventions on academic performance?
    a. Is there a relationship between the characteristics of the goal-setting interventions and their outcomes?
3. How has technology been used to deliver, support, and enhance goal setting in prior studies?

### 4.2 Study 2: Overview and RQs

Building on the findings of the systematic review to inform the design and development of a goal-setting intervention, this study aims to explore the extent to which goal-setting guidance improves the quality of students' goals and how this relates to goal attainment. Furthermore, this study includes metacognitive feedback on the quality of students' goals as part of the SRL cycle. This study adopts a 2x2 factor RCT design, to explore the effects of guidance and feedback.

1. What is the effect of guidance and feedback, individually and combined, on quality of self-set academic goals over time?

### 4.3 Study 3: Overview and RQs

Extending the findings of study 2, with Study 3, we aim to close the gap between improved goal quality and goal attainment by offering guidance and feedback during the performance phase as a bridging support while students work on the goals they have set. Like study 2, this study will follow a 2x2 factor RCT design to explore the effect of guidance and feedback on goal monitoring quality and goal attainment. This guidance takes the form of prompts and scaffolds of the goal monitoring process, with feedback being offered on the quality of the SRL process.

1. What is the effect of prompted goal monitoring and feedback on goal attainment?
2. What is the effect of guidance and feedback, individually and combined, on the content and quality of students goal monitoring over time?

### 4.4 Study 4: Overview and RQs

This study will be conceptualized based on the findings of study 3. The intention is to use this study to explore the final stage of the SRL cycle, reflection, and how guidance and feedback during this phase can improve students' SRL process, and thus, their performance.

## 5 CURRENT STATUS OF WORK

Studies 1 and 2 of this project have been carried out, and we are preparing for the data collection phase of Study 3. For Study 1, we carried out a systematic literature review of goal-setting interventions in higher education. There were three main findings from this study which informed the following studies. Firstly, there is little consensus on how to best design goal-setting interventions for higher education settings, however, on the basis of the examined papers, we selected a goal-setting approach to adapt to use in future studies (Lockspeiser et al., 2013). Secondly, while goal-setting interventions are commonly implemented in higher education settings, these studies often examine the content of students' goals, and rarely test whether the intervention has the desired outcome. And thirdly, while interventions are often delivered digitally, there are very few examples of ICT being used to enhance and adapt the interventions to improve their effectiveness.

With Study 2, we designed and implemented a goal-setting chatbot, testing the effectiveness of the goal-setting intervention alone, and combined with feedback on the quality of students' goal-setting process (Martins van Jaarsveld et al., 2023b). This study showed that goal-setting guidance and feedback were most effective when combined, resulting in significantly higher quality goals. However, increased goal quality was not related to increased goal attainment, highlighting the need for further support during the performance phase of the SRL cycle.

Studies 3 and 4 will focus on using text mining on student interactions with the intervention carried out by the conversational agent to offer real-time, personalized metacognitive feedback during the performance and reflection phases.

## 6 REFERENCES

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, *27*, 1–13. https://doi.org/10.1016/J.IHEDUC.2015.04.007

Heikkinen, S., Saqr, M., Malmberg, J., & Tedre, M. (2022). Supporting self-regulated learning with learning analytics interventions – a systematic literature review. *Education and Information Technologies 2022 28:3*, *28*(3), 3059–3088. https://doi.org/10.1007/S10639-022-11281-4

Hew, K. F., Huang, W., Du, J., & Jia, C. (2022). Using chatbots to support student goal setting and social presence in fully online activities: learner engagement and perceptions. *Journal of Computing in Higher Education*, *35*(1), 40–68. https://doi.org/10.1007/S12528-022-09338-X/

Khosrawi-Rad, B., Rinn, H., Schlimbach, R., Gebbing, P., & Yang, X. (2022). *Conversational Agents in Education: A Systematic Literature Review Review*. https://aisel.aisnet.org/ecis2022_rp/18

Latham, G. P., & Locke, E. A. (2007). New developments in and directions for goal-setting research. *European Psychologist*, *12*(4), 290–300. https://doi.org/10.1027/1016-9040.12.4.290

Lee, H. W., Lim, K. Y., & Grabowski, B. L. (2010). Improving self-regulation, learning strategy use, and achievement with metacognitive feedback. *Educational Technology Research and Development*, *58*(6), 629–648. https://doi.org/10.1007/S11423-010-9153-6/FIGURES/6

Lockspeiser, T. M., Schmitter, P. A., Lane, J. L., Hanson, J. L., & Rosenberg, A. A. (2013). A Validated Rubric for Scoring Learning Goals. *MedEdPORTAL*. https://doi.org/10.15766/MEP_2374-8265.9369

Martins van Jaarsveld, G., Wong, J., Baars, M., Specht, M., & Paas, F. (2023a). Goal Setting Interventions in Higher Education: A Systematic Review. *[Manuscript in Preparation]*.

Martins van Jaarsveld, G., Wong, J., Baars, M., Specht, M., & Paas, F. (2023b). Using a conversational agent to support goal setting amongst higher education students. *[Manuscript in Preparation]*.

McTear, M., Callejas, Z., & Griol, D. (2016). Introducing the Conversational Interface. In *The Conversational Interface: Talking to Smart Devices* (pp. 1–7). Springer International Publishing. https://doi.org/10.1007/978-3-319-32967-3_1

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/FPSYG.2017.00422

Perski, O., Crane, D., Beard, E., & Brown, J. (2019). Does the addition of a supportive chatbot promote user engagement with a smoking cessation app? An experimental study. *Digital Health*, *5*. https://doi.org/10.1177/2055207619880676

Ryan, A. M., Pintrich, P. R., & Midgley, C. (2001). Avoiding seeking help in the classroom: Who and why? *Educational Psychology Review*, *13*(2), 93–114. https://doi.org/10.1023/A:1009013420053

Tsai, Y.-S. (2019). *What is learning analytics?* Society for Learning Analytics Research (SoLAR). https://www.solaresearch.org/about/what-is-learning-analytics/

Uğur, N. G., & Guliz Ugur, N. (2020). Digitalization in Higher Education: A Qualitative Approach. *International Journal of Technology in Education and Science*, *4*(1), 18–25. https://doi.org/10.46328/IJTES.V4I1.24

Wang, G., Wang, Y., & Gai, X. (2021). A Meta-Analysis of the Effects of Mental Contrasting With Implementation Intentions on Goal Attainment. *Frontiers in Psychology*, *0*, 1299. https://doi.org/10.3389/FPSYG.2021.565202

Warriem, J., Thangaraj, A., Mukund, M., & Balaji, B. (2022). Digital interventions at scale. *Communications of the ACM*, *65*(11), 54–57. https://doi.org/10.1145/3550473

Wong, J., Baars, M., Davis, D., Van Der Zee, T., Houben, G. J., & Paas, F. (2019). Supporting Self-Regulated Learning in Online Learning Environments and MOOCs: A Systematic Review. *International Journal of Human–Computer Interaction*, *35*(4–5), 356–373. https://doi.org/10.1080/10447318.2018.1543084

Wong, J., Baars, M., Koning, B. B. de, Zee, T. van der, Davis, D., Khalil, M., Houben, G.-J., & Paas, F. (2019). Educational theories and learning analytics: From data to knowledge. *Utilizing Learning Analytics to Support Study Success*, 3–25. https://doi.org/10.1007/978-3-319-64792-0_1

Zimmerman, B. J., & Schunk, D. H. (2011). Self-Regulated Learning and Performance : An Introduction and an Overview. *Handbook of Self-Regulation of Learning and Performance*, 15–26. https://doi.org/10.4324/9780203839010-4

# Educational design in higher education - transitions to new models

**Alrike Claassen**
University of South Australia
claay037@mymail.unisa.edu.au

**ABSTRACT**: This research investigates instructors' decision-making and the use of LA and AI to inform their educational design process. While LA aims to provide instructors with the information to improve their design practice and the learning environment, instructors face significant challenges with adopting LA in practice. One of the key challenges is that LA systems present the data without alignment with the instructor's design. This research aims to address this challenge by using a phenomenological approach to investigate instructors' use of LA and AI in design practice in the context of higher education. Using self-determination theory as a conceptual foundation for understanding instructors' decision-making process, this research intends to deepen our understanding of instructors' use of digital data to inform their educational design process. Preliminary findings reveal adoption challenges such as fragmented data and overcomplicated dashboards. Future data analysis would explore AI-informed design, and the final stage of the study would involve conceptual mapping with LA and AI experts. The findings will provide actionable insights into how LA and AI can support instructors' design practices.

**Keywords: Learning analytics adoption, LA adoption, generative artificial intelligence, AI adoption, teaching as design, higher education, educational design**

## BACKGROUND

The quality of teaching and design heavily depends on the instructor's ability to engage in reflective practice and consider the needs of students in designing and redesigning learning experiences. During the design phase, instructors determine their pedagogical intent, plan students' learning activities and, based on previous experience, anticipate ways of structuring and scaffolding students' learning (Goodyear, 2015). In the design implementation phase, instructors implement the design and engage with students in various teaching activities, namely sharing content, exchanging knowledge and skills and providing student support (Branch & Kopcha, 2014) (Goodyear, 2015).

At the end of the semester, instructors reflect on their experience and decide on the next iteration based on their reflections and student feedback through formal institutional course evaluation surveys. Instructors' reflective and evaluative processes ideally result in quality course enhancements by responding to student feedback and addressing identified needs.

The decision-making process related to 'teaching as design" is intricate and complicated. In this regard, Self-Determination Theory (SDT) posits that people are motivated to make choices based on three needs: autonomy, relatedness, and competence and that these needs are critical to individual optimal functioning (Ryan & Deci, 2017). SDT is relevant to a teacher's design and teaching practice since the theory proposes that workplace control measures can harm performance (Ryan & Deci, 2017).

Furthermore, SDT advocates that intrinsic motivation can flourish through feedback centred on competence and autonomy in decision-making processes. Building on the principles of SDT, learning analytics feedback, which could provide instructors with pertinent data about students' engagement within learning environments, can potentially increase their intrinsic motivation, especially when instructors possess the autonomy to make decisions based on such information at their discretion.

Studies have shown that LA can support instructors' decision-making and understanding of students' learning experiences that contribute to the educational design and teaching practice evaluation process (Wiley et al., 2020). The trace data obtained in digital learning environments such as a learning management system (LMS) can be seen as digital footprints as students navigate their learning journey. Through the development of interactive dashboards and data visualisations, LA has enabled instructors to gain insights into their students' progress through learning environments (Verbert et al., 2020). Instructors can use LA data to complement previously considered evaluation data, such as their own reflections and student feedback. (Lockyer et al., 2013) instructors illuminated the link between learning analytics, learning design and evaluation, stating that in could use a range of trace data to evaluate their design. Gaining insights from data can only happen when instructors can make sense of the data presented to them, and consequently, turning insights derived from LA into teaching strategies would help close the feedback loop (Clow, 2012). As Leeuwen et al. (2017) noted, instructors often struggle to understand and act on data.

AI has been proliferating, giving instructors access to easy-to-use large language models that could enhance teachers' design practice. However, applying and automating learning environments have also sparked fears among instructors about replacing or adjusting their roles (Zhai et al., 2021). Loeckx (2016) highlighted using AI for personalised learning experiences as an opportunity that could significantly enhance student learning. Instructors tend to move from resistance to overreliance on AI technology (Zhai et al., 2021). In this regard, teachers' perceptions greatly influence the use of AI technologies.

The role of LA and AI in supporting and improving educational design and instructor practice is yet to reach its full potential. Zhai et al. (2021) state that more research investigating the instructor's perspective on AI is needed. Moreover, research is required to identify how LA can enable iterative evaluation, redesign, and evidence-based teaching practice (Mangaroska & Giannakos, 2018). This study will explore how LA can support instructors' decision-making when evaluating and adjusting their educational design or teaching delivery by providing analytics aligned with their pedagogical intent. Drawing on SDT, AI and LA research, the study will examine instructors' decision-making when using LA to inform their educational design and teaching practice. Furthermore, the study will explore how AI can inform the design of instructors' learning events. There is a gap between AI and LA research and adoption, as many instructors rely on their own experiences, perceptions, and institutional data sources to create or revise student learning events. AI has the potential to enhance instructors' design practice, and LA has the potential to offer more nuanced and frequent access to data that can guide design decisions. Although there has been some research on the role of SDT in educational settings, in the context of educational design, this is still an under-researched area. This current research will fill the SDT, educational design, LA and AI research gap by investigating how instructors use LA and or AI to inform their decision-making and how intrinsic drivers influence their decisions.

# RESEARCH GOAL AND QUESTIONS

The present study addresses the research gap by examining how LA can support teachers' decision-making when evaluating and adjusting their educational design or teaching facilitation by providing analytics aligned with their pedagogical intent. The researcher will explore this gap through the following overarching research question:

*How can LA and AI inform instructors' decision-making related to their educational design?*

The researcher will further investigate this question through sub-questions spread over two phases: Phase One focuses on instructors' perceptions and ideals. Phase Two builds on Phase One, where the researcher will consult LA and AI experts and instructors.

## 1.1 Phase One

- ***RQ1:*** How do instructors approach iterative educational design?
- ***RQ2:*** How do instructors use AI or LA in their design, delivery and evaluation?
- ***RQ3:*** How would instructors prefer to use data types or generative AI to evaluate course design, delivery and evaluation in the future?
- ***RQ4:*** How do instructors prioritise and make decisions based on LA or generative AI?

## 1.2 Phase Two

- ***RQ5:*** How can LA or AI inform instructors' course design, delivery and evaluation?
- ***RQ6:*** What data and steps can be established to enable LA and generative AI for iterative design?

# RESEARCH METHODOLOGY

This study explores how LA and generative AI can support instructors' decision-making when evaluating and adjusting their educational design or teaching practice by providing information aligned with their pedagogical intent. In congruence with the aim of the study, the researcher will adopt an interpretive stance (Denzin & Lincoln, 2018). Therefore, the qualitative nature of the research aim will determine the research design.

## 1.3 Phase One: LA and AI focus groups

Parts 1A and B aim to explore the realities of redesign and how instructors in higher education are utilising LA and AI by describing their current educational design practices. In Part 1A, I unpacked how instructors use LA to inform their design. In this phase, a phenomenological approach was utilised, seeing that the literature in this regard is still scarce Bennett et al. (2017), and to gain insight into the realities of redesign, articulation of pedagogical intent, and the data instructors would like to use in the future or already make use of when making evidence-based decisions. The researcher conducted focus groups to elicit conversation. The focus group participants were selected through purposeful sampling. The recruitment happened by e-mail, and selection criteria were established for participant inclusion, allowing only instructors with course coordination experience and experience with learning analytics to be included. Each focus group included between 2 to 6 participants. A semi-structured interview protocol guided the discussions in the focus groups. During Phase 1B, the researcher recruited the same participants as in Phase 1A to better understand how instructors use AI to enhance or inform their educational design. The requirement was that instructors should already use AI in order to be included in focus group participation. In phases 1A and 1B, the researcher will identify gaps in knowledge, challenges and solutions as perceived by instructors. The findings will be collated and synthesised. The synthesised results will highlight the challenges instructors perceive and their associated practices to inform Phase Two of the study.

### 1.4 Phase Two: Conceptual mapping

The aim of Phase Two is to explore the potential for LA in enhancing iterative educational design, focusing on how LA can support instructors in their educational design, delivery, and evaluation, as well as determining the data and steps required to enhance the quality of their designs. The findings from Phase 1 identify the specific areas where instructors can deploy AI to support the iterative educational redesign process. To address the identified knowledge gaps, phase 2 will employ conceptual mapping, drawing on the approach outlined (Colvin et al., 2016). The authors used a concept mapping methodology in this study to expand and critique current practice. The researcher will select the expert panel from expert researchers in learning analytics and educational or learning design. The researcher will identify expert researchers through a review of past Society of Learning Analytics Research (SOLAR) publications, including international conference proceedings and workshops and the Journal for Learning Analytics. The researcher will provide the findings from Phase One and a brainstorming prompt to the expert panel for the concept mapping. The brainstorming prompt will be: *How can learning analytics and artificial intelligence inform instructors' design, course delivery and evaluation of courses?* Responses will be analysed using the Thematic Analysis Process (Denzin & Lincoln, 2018). The thematic analysis will start with the participant's response and create a list of statements from there. This way, the researcher will merge corresponding responses into one statement, remove redundancies and make a list of statements. Following the extraction of themes, the identified statements will be sent back to the expert panel for the second round, in which they will rank the identified statements by their importance and complexity of implementation. The results of phase 2 will provide researchers with actionable insights into how instructors can use LA to support educational design practice.

## PRELIMINARY FINDINGS: LA-INFORMED DECISION-MAKING

Based on instructors' present practices, some key findings on how they currently use LA to inform their design include learning resources and assessment data. The adoption challenges experienced by instructors included a need for more professional development and issues of fragmented data presentation and frustrating dashboards. Moreover, there is a lack of professional development and time to allow better analytics integration in instructors' design practices. As expected, instructors were very positive about AI mitigating these adoption challenges. Instructors suggested predictive analytics to identify at-risk students, shared dashboards between students and instructors, improved dashboards and visualisations, student profiles and artificial intelligence data interpretation. Based on these findings, LA developers could redesign LA systems based on three suggestions: LA that provides design-specific AI-interpreted feedback. LA dashboards that enable customisation and improved visualisation and LA that support social interaction and allow early intervention.

# CONTRIBUTION

The PhD study focuses on how LA and AI can shape instructors' decision-making from the theoretical basis of self-determination theory. Firstly, the research will contribute to current understanding by elucidating instructors' current design practices informed by AI and LA. Secondly, the research will unpack opportunities and challenges as perceived by instructors. Thirdly, the researcher will create a conceptual framework of LA and AI aligned with instructors' educational design needs.

# REFERENCES

Bennett, S., Agostinho, S., & Lockyer, L. (2017). The process of designing for learning: Understanding university teachers' design work. Educational Technology Research and Development, 65(1), 125–145. https://doi.org/10.1007/s11423-016-9469-y

Branch, R. M., & Kopcha, T. J. (2014). Instructional Design Models. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), Handbook of Research on Educational Communications and Technology (pp. 77–87). Springer New York. https://doi.org/10.1007/978-1-4614-3185-5_7

Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. 134–138.

Colvin, C., Rogers, T., Wade, A., Dawson, S., Gasevic, D., Buckingham Shum, S., Nelson, K., Alexander, S., Lockyer, L., & Kennedy, G. (2016). Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement.

Denzin, N. K., & Lincoln, Y. S. (2018). The SAGE handbook of qualitative research / edited by Norman K. Denzin, University of Illinois, Yvonna S. Lincoln, Texas A&M University (Fifth edition.). Thousand Oaks, California : SAGE.

Goodyear, P. (2015). Teaching as design. Herdsa Review of Higher Education, 2(2), 27–50.

Leeuwen, A. van, Wermeskerken, M. van, Erkens, G., & Rummel, N. (2017). Measuring teacher sense making strategies of learning analytics: A case study. Learning (Abingdon (England)), 3(1), 42–58. https://doi.org/10.1080/23735082.2017.1284252

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing Pedagogical Action:Aligning Learning Analytics With Learning Design. American Behavioral Scientist, 57(10), 1439–1459. https://doi.org/10.1177/0002764213479367

Loeckx, J. (2016). Blurring boundaries in education: Context and impact of MOOCs. International Review of Research in Open and Distributed Learning, 17(3), 92–121.

Mangaroska, K., & Giannakos, M. N. (2018). Learning analytics for learning design: A systematic literature review of analytics-driven design to enhance learning. IEEE Transactions on Learning Technologies, 1–1. https://doi.org/10.1109/TLT.2018.2868673

Ryan, R. M., & Deci, E. L. (2017). Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. New York: Guilford Publications. https://doi.org/10.1521/978.14625/28806

Verbert, K., Ochoa, X., Croon, R. D., Dourado, R. A., & Laet, T. D. (2020). Learning analytics dashboards: The past, the present and the future. 35–40. https://doi.org/10.1145/3375462.3375504

Wiley, K. J., Dimitriadis, Y., Bradford, A., & Linn, M. C. (2020). From theory to action: Developing and evaluating learning analytics for learning design. Proceedings of the Tenth International

Conference on Learning Analytics & Knowledge, 569–578. https://doi.org/10.1145/3375462.3375540

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J.-B., Yuan, J., & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. Complexity, 2021, 8812542. https://doi.org/10.1155/2021/8812542

# Privacy and Data Protection for Trustworthy Learning Analytics in Higher Education

**Qinyi Liu**
Centre for the Science of Learning and Technology (SLATE), University of Bergen, Bergen, Norway
qinyi.liu@uib.no

**ABSTRACT**: This PhD project explores the privacy and data protection issues that exist in learning analytics, how they can be addressed, and to establish trustworthy learning analytics by better handling privacy and data protection of learning analytics. The PhD project encompasses three planned studies. The first study is a systematic literature review that summarizes the identified privacy and data protection issues, as well as previously proposed solutions in the field. The second study addresses selected privacy issues in learning analytics through a generative AI approach and provides quantitative metrics to easily interpret the results. The third study combines the privacy and data protection issues identified in the first study and the synthetic data generation algorithms in the second study into a toolkit that can be customized and used by the broader learning analytics community. With the results of the three research studies, this PhD project bridges many of the shortcomings of previous solutions and uses a state-of-the-art solution to better address learning analytics privacy issues. Additionally, this dissertation aims to develop a privacy-preserving learning analytics toolkit that can enable a scalable learning analytics while encouraging data sharing and open science practices, thus fostering a more trustworthy learning analytics environment.

**Keywords**: Learning analytics, privacy, data protection, trustworthy learning analytics, synthetic data generation, differential privacy

## 1 INTRODUCTION AND MOTIVATION

In recent years, the field of learning analytics (LA) has made significant progress with the continuous development of AI technologies, but at the same time, its shortcomings in terms of privacy and data protection have become more serious, hindering its further expansion (Joksimović et al., 2021). There are three main reasons why privacy and data protection are so important for LA and its practices. One is because data usability and privacy are inversely related, and more data usability requires compromising on privacy, and vice versa (Prinsloo et al., 2022). Therefore, if privacy issues are not properly addressed and balanced in LA, it will greatly affect the results of LA and thus the credibility of the field. The second reason is that privacy and data protection are important pillars of trustworthy LA. The rationale for trustworthy LA is that LA can only reach its maximum potential if trust is established in the process of development, deployment, and use (Thiebes, Lins & Sunyaev, 2020). However, to build trust among stakeholders, and stakeholder trust in LA systems, privacy and data protection issues must be properly addressed. The third reason is that the current laws such as the General Data Protection Regulation (GDPR) in Europe set a list of rules that must be followed to protect the rights of collecting, processing, using, and storing of individual data. Violating the GDPR articles risks legal repercussions that could lead to significant financial costs and serious consequences to institutions such as higher education (Kitto & Knight, 2019). Moreover, at the legal level, the introduction of the GDPR presents many new challenges and opportunities for LA research and practice: anonymizing data, limiting personalized interventions in LA (Karunaratne, 2021). In addition, new trends such as the widespread use of social media in learning and multimodal learning have increased the difficulty of privacy issues (Joksimovic et al., 2021). The current solutions that have been proposed are difficult to cope with the ever-complex privacy and data protection challenges in LA (Liu & Khalil, 2023). Driven by the above reasons, this Ph.D. project is themed to focus on privacy and data protection issues in LA. By identifying and addressing the increasingly complex and important privacy and data protection issues in LA, this Ph.D. impact will not only support trustworthiness of LA but also help scale the field.

## 2 BACKGROUND REVIEW

This background review section is divided into two parts, one of which will be a discussion of the definition of privacy and data protection in LA, and the second will be an elaboration of the prior research on privacy and data protection in LA. I will go through the background review and elaborate on the research gaps that exist in this area.

### 2.1 Definition of privacy and data protection

Privacy and data protection are complex and multidimensional concepts, and there is no single theoretical framework applicable to all scenarios (Page & Wisniewski, 2022). There have been several studies that attempt to explore the concept of privacy in LA, but there is not yet a clear as well as agreed definition of privacy that is recognized by the LA community (Liu & Khalil, 2023). In this Ph.D. project, the definition of privacy will be summarized from previous LA findings as students having control over their data, and not disclosing personal information throughout the data collection, analysis, or reporting process. With respect to data protection, this Ph.D. project adopt the definitions of data protection as stipulated by the data protection laws of the United Kingdom (UK) and the European Union (EU): (1) all activities related to personal data, such as collection, processing, storage, transmission and deletion, must be used for specific and explicit purposes in accordance with the principles of fairness and transparency, (2) measures must be taken during data processing to ensure that personal data are handled in a secure manner and (3) data must be kept up to date during this process and saved only for as long as necessary (Data Protection Act, 2018; Kuner et al., 2020).

### 2.2 Prior research on privacy and data protection in learning analytics

Although the introduction section shows the continuing importance of privacy and data protection for LA, research in this area is still at a relatively early stage. This is reflected in two aspects: first, there is a lack of a systematic overview of what privacy and data protection issues arise at different phases of LA (e.g., data collection, data sharing). The other aspect is that existing privacy protection solutions are limited, ad hoc, and cumbersome, and do not adequately address privacy and data protection issues in LA (Joksimović et al., 2021).

First, in exploring privacy and data protection issues in LA, Drachsler and Greller (2016) discuss existing problems, emphasizing power relations, data/user exploitation, data ownership, and several other aspects. Other studies discuss privacy and data protection issues in LA in a decentralized manner, such as Torre et al.'s (2020) study on the risks of remote computing of sensitive data in the context of digital education. However, previous studies lacked systematic answers on what privacy and data protection issues are involved in the different phases of LA (data collection, data analytics, data sharing and publication). As AI technologies continue to evolve, and online learning and education expand, previously unnoticed privacy and data protection challenges within the domain of LA may come to the forefront. There is an urgent need for a new, systematic mapping of privacy and data protection issues in LA.

Second, existing solutions are not sufficient to address privacy and data protection issues in LA in the context of growing challenges. Existing solutions can be categorized into legal and framework-based solutions, technical solutions, and combined solutions (Liu & Khalil, 2023). Combined solutions are not discussed for the time being as there are fewer of them and those that have been proposed have not yet been finalized. As for technical solutions, very well-known examples include the MOOC Replication Framework (MORF) created by Hutt et al. (2022), which allows researchers to use data without direct access. However, this approach still has its limitations, there are still many privacy challenges as the MORF still anticipates sharing data in its raw data format (Joksimovic et al., 2022) and the tool can only be used in a MOOC environment and lacks scalability. In addition, some of the previous privacy-preserving technical solutions, may no longer be effective, e.g., Yacobson et al. (2021) found that the use of unsupervised machine learning can re-identify LA data that has been anonymized. On the other hand, other technical solutions suffer some drawbacks, such as trade-offs between privacy and data utility (Prinsloo et al., 2022), and are costly, ambiguous, and not easily scalable (Prinsloo et al., 2022). As for legal and framework-based solutions, representative examples include Drachsler & Greller (2016), who, based on an in-depth study of legal texts and a literature review, propose an eight-point checklist to help educational institutions decipher privacy and ethical issues in LA. Other examples include Hoel and colleagues' proposal for a separate research area on the impact of legal frameworks on LA (Hoel et al., 2017). While these legal and framework-based solutions can provide some guidance on the overall design of learning analytics, they tend to be conceptual and lack clear actionable guidelines and evidence of application (Marshall et al., 2022; Liu & Khalil, 2023). Therefore, the background review found a necessary need for a solution that can provide robust privacy protection in the face of rapid AI development.

## 3 Research goals and questions

According to the motivation demonstrated in the introduction section and the research gaps that have been identified by the previous background review section, the overarching aim driving this Ph.D. research project is: What are the privacy and data protection issues surrounding LA in latest development, how to address them, and establish a more trustworthy learning analytics ecosystem?

This question will be explored through several more granular inquiries related to privacy-driven and trustworthy learning analytics, sub-question for each study is summarized here: (1) What are the identified privacy and data protection issues throughout the LA process, from data collection to data reporting? (2) How to use state-of-art approaches to address the emerging complex privacy issues such as insufficient anonymization in LA? (3) How can the proposed solutions overcome the shortcomings of previous solutions, maintain the balance of utility and privacy, be easy to use and extend, and be able to be customized for different scenarios?

# 4 Studies and methods

## 4.1 Study 1: understanding privacy and data protection issues in LA using systematic review (has been published)

To answer the first RQ, in my first doctoral study, I aim to conduct a systematic review of the literature, with LA, privacy, and data protection as key entry points. The purpose of the systematic review is to investigate what privacy and data protection issues there are in LA and how previous solutions addressed privacy issues and construct trustworthy learning analytics. The first article of the PhD project provides an important foundation for the next phase of solution development by mapping privacy and data protection issues around learning analytics, as well as summarizing the field's efforts to build trustworthy learning analytics.

The research questions that the first study attempts to answer are (1) What are the identified privacy and data protection issues throughout the LA process, from data collection to data reporting? (2) How do stakeholders from various backgrounds view privacy and data protection issues in LA similarly and differently? (3) How has previous research attempted to address the privacy and data protection issues identified in LA?

## 4.2 Study 2: scaling while privacy preserving: synthetic data generation in learning analytics (has been submitted)

To answer the second RQ, based on the result of the first study I use a state-of-the-art approach (i.e. generative AI and differential privacy) to address the increasing complexity of privacy issues in this area. Specifically, the results of my first study (see section 4.1) found that the existing solutions are difficult to address insufficient anonymization, data misuse and sensitive data local storage computing problems well, and most of the solutions lack the support of practical evidence. With such motivation, the second study will use generative algorithms to generate common data types in LA, such as tabular data, time series data. Synthetic data can perform a variety of functions. For example, it can be used to replace hard-to-obtain samples and reduce the cost of data collection (Turing, 2023). In terms of privacy, synthetic data can be used as a substitute for sensitive data that cannot be migrated, thus satisfying privacy regulations by GDPR. In addition, in cases where data sharing is required, synthetic data can be used as a substitute for the real data to protect the privacy of the real data, thus addressing the issues of insufficient anonymization and data misuse from a different angle. In response to the lack of practical evidence for the previous LA privacy and data protection solutions, I will evaluate and test synthetic data in terms of three dimensions: resemblance (how similar the synthetic data is to the real data), utility (how synthetic data is performing in terms of data analytics and machine learning), and privacy (whether synthetic data leaks information about real data and how effective it is in countering attacks).

In addition to using synthetic data generation for supporting privacy protection, this paper will also use a differential privacy approach. Differential privacy provides higher security and privacy than previous de-identification techniques and retains the primary dataset at the data custodian, allowing querying and use of the data, but not migration of the data (Dyda et al., 2021). Differential privacy and synthetic data generation will be used in combination to compensate for each other's strengths and weaknesses. For example, synthetic data can be helpful to small datasets to scale while providing privacy protection, whereas data synthesis using large datasets (e.g., more than a million rows) would be computationally intensive. In the case of large datasets, differential privacy safeguards large datasets effectively due to its noise addition method, making it suitable for privacy protection in extensive datasets (Dyda et al., 2021). Finally, synthetic generation algorithms with added differential privacy features provide higher privacy protection, which can provide more reliable support for high privacy demanding LA scenarios.

The research questions that the second paper attempts to answer are (1) How to make a comprehensive assessment of synthetic data, including the dimensions of resemblance, utility, and privacy, in the LA field? (2) To what extent the use of synthetic data can improve the privacy aspects in LA predictive modeling while maintaining the model's predictive performance? (3) How to customize the generation of synthetic data to suit different scenarios in LA predictive modeling?

### 4.3 Study 3: toolbox for synthetic data generation and privacy mapping dashboard

Grounded in the outcomes of the previous two studies, in Study 3, I will develop a toolkit. This toolkit will translate the results of the first two studies on privacy and data protection into an easier, less-threshold approach for the wider LA community. The toolkit provides two contributions (1) Based on the results of the first article, it will provide a mapping tree of the privacy and data protection issues in the different phases of LAs, (2) Based on the results of the second paper, it will allow users to use the toolkit according to their individual needs (e.g., different needs for data utility and privacy) and provide state-of-the-art differential privacy services for large datasets. To better address privacy and data protection issues in LA, the toolkit should be able to overcome the opacity, cost and non-transparency issues mentioned in previous literature (Prinsloo et al., 2022). Therefore, Study 3 will conduct user testing to get qualitative feedback from users to improve the toolkit to overcome the shortcomings of previous solutions.

The research questions that the third study aims to answer are (1) How can this toolkit provide clear privacy and data protection mapping and privacy-protected data generation in LA that can be customized for different scenarios? (2) To what extent can an easy-to-use, low-threshold, and easily scale privacy-enhancing toolkit facilitate the resolution of privacy and data protection issues in LA?

## 5 PRELIMINARY RESULTS

First study about the systematic review on privacy and data protection issues in LA has been published in the British Journal of Educational Technology (See detail in Liu and Khalil, 2023). This article identifies eight privacy and data protection issues at various phases in LA, summarizes two similarities and three differences in stakeholder perceptions of privacy and data protection, and divides previous attempts to address privacy and data protection issues into three categories and evaluates them. Specifically, this paper finds that previously proposed solutions are not very effective in terms of privacy issues such as insufficient anonymization, data misuse, local storage, and calculation of sensitive data in the different LA phases. Additionally, the majority of the solutions lack the support of practical evidence.

Study 2 has been partially carried out and the work of synthetic tabular data generation has been completed and the synthetic data has been evaluated in terms of three dimensions, namely, resemblance, utility, and privacy, using different metrics. Preliminary results show that synthetic data can be improved in terms of privacy while maintaining similar utility as real data. The next work on synthetic time series data generation will be conducted and similarly evaluated on all three dimensions. As for Study 3, the domain name (http://lasd.ai) has already been created, and the development of the toolkit will begin after the completion of Study 2.

## REFERENCES


Data Protection Act, Gov.uk. (2018). https://www.gov.uk/data-protection

Drachsler, H., & Greller, W. (2016). Privacy and analytics: it's a DELICATE issue a checklist for trusted learning analytics. ACM Digital Library; Association for Computing Machinery. https://doi.org/10.1145/2883851.2883893

Dyda, A., Purcell, M., Curtis, S., Field, E., Pillai, P., Ricardo, K., Weng, H., Moore, J. C., Hewett, M., Williams, G., & Lau, C. L. (2021). Differential privacy for public health data: An innovative tool to optimize information sharing while protecting data confidentiality. *Patterns, 2*(12), 100366. https://doi.org/10.1016/j.patter.2021.100366

Joksimović, S., Marshall, R., Rakotoarivelo, T., Ladjal, D., Zhan, C., & Pardo, A. (2021). Privacy-Driven Learning Analytics. In Manage Your Own Learning Analytics (pp. 1–22). Springer. https://doi.org/10.1007/978-3-030-86316-6_1

Karunaratne, T. (2021). For Learning Analytics to Be Sustainable under GDPR—Consequences and Way Forward. *Sustainability, 13*(20), 11524. https://doi.org/10.3390/su132011524

Kitto, K., & Knight, S. (2019). Practical ethics for building learning analytics. *British Journal of Educational Technology, 50*(6), 2855–2870. https://doi.org/10.1111/bjet.12868

Kuner, C., Bygrave, L. A., & Docksey, C. (2020). The EU General Data Protection Regulation (GDPR). Oxford University Press.

Liu, Q., & Khalil, M. (2023). Understanding privacy and data protection issues in learning analytics using a systematic review. *British Journal of Educational Technology*, 54, 1715–1747. https://doi.org/10.1111/bjet.13388

Marshall, R., Pardo, A., Smith, D., & Watson, T. (2022). Implementing next generation privacy and ethics research in education technology. *British Journal of Educational Technology*, 53. https://doi.org/10.1111/bjet.13224

Page, X., & Wisniewski, P. (2022). *Modern socio-technical perspectives on privacy (*B. P. Knijnenburg, X. Page, P. Wisniewski, H. R. Lipford, N. Proferes, & J. Romano, Eds.). Springer International Publishing. https://doi.org/10.1007/978-3-030-82786-1

Prinsloo, P., Slade, S., & Khalil, M. (2022). The answer is (not only) technological: Considering student data privacy in learning analytics. *British Journal of Educational Technology*, 53, 876–893. https://doi.org/10.1111/bjet.13216

Torre, M. V., Tan, E., & Hauff, C. (2020). edX log data analysis made easy. Proceedings of the Tenth International Conference on Learning Analytics & Knowledge. https://doi.org/10.1145/3375462.3375510

Thiebes, S., Lins, S., & Sunyaev, A. (2020). Trustworthy artificial intelligence. *Electronic Markets* 31, 447–464. https://doi.org/10.1007/s12525-020-00441-4

Yacobson, E., Fuhrman, O., Hershkovitz, S., & Alexandron, G. (2021). De-identification is Insufficient to Protect Student Privacy, or – What Can a Field Trip Reveal? *Journal of Learning Analytics, 8*(2), 83–92. https://doi.org/10.18608/jla.2021.7353

# Culturally Aware Human-Centred Design of a Cognitive-affective State-based Online Learning Analytics Tool

**Author(s): Victor Manuel Corza Vargas**

Graduate Program in Computer Science and Engineering, UNAM, México.
victorcorza@comunidad.unam.mx

**ABSTRACT**: *Online learning environments offer unmatched flexibility and accessibility, but there are several challenges that need to be resolved. Compared to a physical classroom, where teachers are aware of the emotional atmosphere, allowing them to adjust their teaching, when necessary, in synchronous online environments replicating this situation is complicated since the teacher has a limited view of students. In addition, physical classrooms allow learners to be aware of each other's emotional states, resulting in richer interaction and collaboration; this type of interaction is not possible in synchronous online settings. The concept of social translucence, which means rendering learners' affective states visible to teachers and other learners through technology, offers an alternative for enhancing awareness of affective states through learning analytics (LA) tools. Current advances in AI and other areas of computer technology are enabling systems to monitor affective states. However, as these types of LA tools become more widespread, several cultural, ethical, and privacy concerns arise. Human-centred design can be an alternative to address these concerns by eliciting the perspectives of different stakeholders to include in the design of LA tools, making them more appropriate to the real needs of synchronous online learning.*

**Keywords**: Human-centeredness, learning analytics, cognitive-affective states, emotions, ethics, computer vision.

## 1 INTRODUCTION

Affect, in psychology, refers to the underlying experience of feeling, emotion, attachment, or mood (Hogg, M. A., & Abrams, D., 2007). Ekman and Friesen established that emotions are a type of affect with rapid onset, short duration, spontaneous occurrence, automatic evaluation, and coherence between responses (Ekman and Friesen, 1992). Additionally, in the same study they identified six basic emotions: happiness, sadness, fear, anger, surprise, and disgust. Living in a different spectrum cognition refers to all activities and processes concerned with the acquisition, storage, retrieval, and processing of information (Byrne, R. W., et. al., 2019).

These definitions provide the fundamentals for a broader array of states that influence cognition and deep learning in the context of learning computer environments; these states include boredom, engagement, confusion, frustration, delight, and surprise (Calvo & D'Mello, 2011). These states are generally referred to as cognitive-affective states (CASs) because they have significant cognitive and affective components in the context of learning. In (D'Mello, S., & Graesser, A. 2011), these states were placed within Russell's Core Affect framework (Russell, J.A., 2003), which juxtaposes valence (pleasure to displeasure) against arousal (activation to deactivation), positioning basic emotions based on their arousal and valence values, see Figure 1. This framework establishes an effective bridge between basic emotions and cognitive-affective states (CASs). Emotions have a significant impact on learning by affecting the cognitive processes of students. Conversely, the cognitive processes of a student contribute to their emotional responses by influencing the way they think, perceive, and process information. Due to these intricate relationships, it is advisable to design learning analytics tools that consider not only the affective states but also the cognitive states of students. Such considerations can lead to increased usage and acceptance of an LA tool (Baker, R. S., 2010).

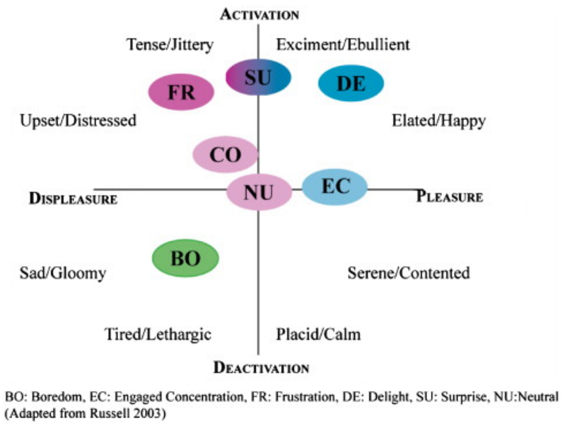



**Figure 1: CASs located in Russell's core affect framework**

On the other hand, human-centred design aims to elicit students and teachers' preferences on several technical and non-technical design aspects (Siemens, G., 2013). The increasing spread of emotion aware LA tools is generating several concerns related to data privacy and ethics, such as information leakage, students profiling, and inadequate use of data. Additionally, these tools, along with the datasets often employed to train these models, frequently overlook broader contextual factors such as cultural differences, even though these are crucial in understanding emotions (Katirai, A., 2023). These human-centred considerations are far from being fully integrated into the design of emotion recognition-based LA systems for synchronous online learning contexts.

## 2 Goals of the research and research questions

The main goal of this research is to attain emotional translucence in synchronous online classes to mimic a physical classroom experience where teachers play a key role in continuously monitoring perceived students' emotional states to tailor learning activities accordingly to foster an emotional climate conducive to students' learning. Moreover, students' awareness of each others' emotional states can also positively contribute to the emotional regulation of the class. Emotional translucence entails rendering CASs visible to both teachers and students through a minimally intrusive LA tool. Additionally in order to integrate teachers and students' opinions in the LA tool design process, I propose an elicitation process to gather their preferences on various areas of design.

This project focuses on three research questions: **RQ1**: What are the students' and teachers' perspectives on the practical, ethical and privacy implications of modelling and visualising student's CASs during an synchronous online class? In RQ1, perspectives are defined as the stakeholders' preferences, expectations, and concerns, based on their personal experience of participating in or conducting synchronous online classes; **RQ2**: To what extent do CASs impact learning and what is the effectiveness of modelling CASs using a combination of facial expressions and Russel's Core Affect framework?; In RQ2, effectiveness stands for how accurate the proposed method is to recognize CASs in synchronous online learning settings. **RQ3**: How can a human-centred approach contribute to addressing the cultural concerns related to a CAS-based LA tool in the synchronous online learning context?

# 3 CURRENT KNOWLEDGE OF THE PROBLEM DOMAIN AND STATE OF EXISTING SOLUTIONS

## 3.1 Facial emotion recognition (FER)

Convolutional Neural Networks (CNNs) can extract features efficiently from a collection of examples. Several architectures based on CNNs have been proposed, for instance, VGG, RESNET, INCEPTION, among many others. A recent evolution in Neural Networks for Computer Vision is the use of Vision Transformers (ViT) (Khan, S., et. al., (2021), where images are treated as sequences of elements, modelling feature maps as token vectors.

## 3.2 Cognitive-affective states modelling

Using different approaches, systems have been proposed to model students' CASs in educational environments, several of them use facial emotion recognition as a subjacent technology for CASs recognition. For instance, (Gupta, S., et. al., 2023) introduced a deep learning-based method using facial emotion recognition to real-time detect online learners' engagement. In a similar vein, (Dawood, A., et. al., 2018) modelled students' affective-cognitive states, such as confidence, uncertainty, engagement, anxiety, and boredom, solely using a webcam. Additionally (Yan, F., et. al., 2022) employed facial expression recognition, speech analysis and CNNs techniques to visualise the emotional dynamics of individual students in an online class by mapping the low-level data with arousal and valence.

## 3.3 Databases in Facial Expression Recognition

To train a Facial Expression Recognition (FER) model, sufficient training databases are required that include as many variations of ethnicities and environments as possible. Because scenes in reality are complex and changeable due to multiple factors such as different backgrounds, occlusions, and luminosities, FER models rely on uncontrolled databases such as FER2013(Karan, K. V., et. al., 2022) and RAF-DB(Li, S., Deng, W., & Du, J., 2017).

## 3.4 Culturally Aware Human-Centred Design

The spread of emotion-aware LA systems is leading to an emerging emphasis on incorporating the views and experiences of learners and teachers into LA systems design and functionality(Dimitriadis, Y., et. al., 2021). This push acknowledges the intricate design elements that stem from students' and teachers' real-world experiences(Martinez-Maldonado, R. 2023). Additionally, there's a rising awareness of the importance of considering broader contextual and cultural dimensions, ensuring that LA systems are sensitive to diverse student backgrounds (Viberg, O., et. al. 2023). Datasets for developing models to detect emotions can be skewed towards specific demographics in terms of cultural representation. For example, the RAF-DB dataset (Li, S., et. al., 2017), sourced from diverse ages, genders, and races, may not fully encompass cultural particularities. Conversely, the JAFFE dataset (Lyons, M., et.al., 1998), short for Japanese Female Facial Expression, offers a regional perspective, showcasing facial expressions of Japanese females, underscoring the need to consider region-specific facial features and cultural indicators.

# 4 Discussion on Novelty and Expected Impact

In the context of CAS-based LA tools this thesis aims at going beyond previous work by proposing a novel approach to model CASs by combining Russel's Core Affect Framework with state of the art deep neural networks such as vision transformers (ViT's) and regional database creation for dramatically improving accuracy on facial expression recognition in the context of higher education in Mexico. On the other hand, this thesis advances the integration of human-centred considerations in three crucial aspects: data privacy, ethics and cultural concerns through accompanying the development of a CAS-based LA tool with a comprehensive elicitation process, aimed at gathering student and teacher experiences and perspectives.

# 5 Methodology

## 5.1 Human Centred Design

In this project I'm following two human-centred frameworks: [1] design thinking and [2] translucent learning analytics (Martinez-Maldonado, R. et al., 2022). To gather the opinions of students and teachers, structured, individual interviews will be carried out to enable comprehensive elicitation about: a) students' perception of the emotions that influence their performance in synchronous online classes; b) practical approaches for visualising students' CASs during synchronous online classes; and c) ethical, privacy, and cultural concerns about modelling students' CASs and sharing them with their teachers and other students. Subsequently, this interview structure will be adapted for elicitation with teachers. A blend of open questions, interactive activities and drawing activities will be employed. For open ended questions an inductive approach will be applied to guide the analysis in order to identify emerging themes. A deductive approach will be applied for interactive activities, such as matching, ranking and expressing agreement/disagreement. For drawing activities an inductive approach will be applied. These activities will pose students to generate a variety of creative visual representations accompanied by verbal explanations. An exploratory review on both will enable the identification and grouping of similar ideas to create a set of preferences for CAS-based LA tool design.

## 5.2 Regional database

A regional database that highlights facial expressions of students aged 19 to 26 enrolled in a higher education institution in Mexico will be generated. This dataset will enhance the predictability of the chosen deep learning model for FER in section 5.3, via supplementary fine-tuning procedures to incorporate specific regional morphological characteristics of Mexican students divided by gender.

## 5.3 Cognitive-affective states modelling

FER will be addressed with a transfer learning approach, by using deep neural networks pre-trained with the ImageNet dataset. These networks will be trained with facial expression databases, like FER2013 and RAF-DB. Data augmentation will be applied to the training sets of FER2013 and RAF-DB datasets to prevent overfitting. All models will be fine-tuned using Pytorch deep learning framework. All tasks comprising FER will be developed using the python programming language.

Several basic emotions gathered during a specific time period will act as components to model a CAS by increasing or decreasing arousal or valence values and depicting these changes in a two dimensional plane, to assign the CAS according to the final position. Therefore I will be utilising a Facial Emotion Recognition (FER) as a bridge for CAS modelling. Additionally, other facial expressions will be considered to offer complementary evidence of some CASs, for example drowsiness and yawning, which are clear signs of boredom or tiredness (Busari A. O., 2018).

## 6 Current status

To identify the students preferences on the design process of an LA tool based on CASs, an elicitation process with students was conducted. Nine 5th-semester Computer Science students from the Faculty of Sciences at UNAM participated in the study. Their average age is 22.5 years (std. dev. = 1.38). Of these, 8 identified as male and 1 as female. The findings of the data analysis were reported in depth in the paper "Students' Perspectives on the Ethical, Privacy, Cultural and Design Implications of Modelling and Visualising their Cognitive-Affective States in Online Classes", which was submitted for LAK 24. In this paper, it was found that students envisaged solutions that incorporated a variety of data sources to aid emotion recognition, with significant emphasis on facial recognition. On the other hand, students predominantly conceptualised dashboards designs to visually represent data.

For addressing FER, several tasks for facial expression recognition had been performed: a) An accurate system for identifying ROIs (Face Detection) and distinguishing individual faces (Face Recognition) has been developed; b) Both FER2013 and RAF-DB datasets have been thoroughly preprocessed and augmented; c) 12 experiments have been conducted on the FER2013 dataset using various deep learning networks. Table 2 in the Appendix briefly describes the most relevant experiments that were carried out.

Finally with the students' approval through an informed consent, recordings of synchronous online sessions from two different groups of the subject 'Database Fundamentals' at Faculty of Sciences were made, Table 1 in the Appendix shows the breakout of recorded sessions. These video recordings include 26 different students that have given their consent for their facial expressions to be included in a regional database of higher education students at UNAM, Mexico. These recordings provide the foundational material for real-time processing tasks, regional database creation and LA tool testing by simulating synchronous online classes.

## REFERENCES

Hogg, M. A., & Abrams, D. (2007). *Social cognition and attitudes.* Pearson Education.

Ekman, P. (1992). *An argument for basic emotions.* Cognition & emotion, 6(3-4), 169-200.

Calvo, R. A., & D'Mello, S. K. (Eds.). (2011). *New perspectives on affect and learning technologies* (Vol. 3). Springer Science & Business Media.

Bayne, T., Brainard, D., Byrne, R. W., Chittka, L., Clayton, N., Heyes, C., ... & Webb, B. (2019). What is cognition?. Current Biology, 29(13), R608-R615.

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.

D'Mello, S. (2017). Emotional learning analytics. Handbook of learning analytics, 115.

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2021). Transformers in vision: A survey. *ACM computing surveys (CSUR)*.

Katirai, A. (2023). Ethical considerations in emotion recognition technologies: a review of the literature. AI and Ethics, 1-22.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. American Behavioral Scientist, 57(10), 1380-1400.

Yan, F., Wu, N., Iliyasu, A. M., Kawamoto, K., & Hirota, K. (2022). Framework for identifying and visualising emotional atmosphere in online learning environments in the COVID-19 Era. Applied Intelligence, 52(8), 9406-9422.

Gupta, S., Kumar, P., & Tekchandani, R. K. (2023). Facial emotion recognition based real-time learner engagement detection system in online learning context using deep learning models. Multimedia Tools and Applications, 82(8), 11365-11394.

Dawood, A., Turner, S., & Perepa, P. (2018). Affective computational model to extract natural affective states of students with Asperger syndrome (AS) in computer-based learning environment. IEEE Access, 6, 67026-67034.

Karan, K. V., Bahel, V., Ranjana, R., & Subha, T. (2022). Transfer learning approach for analyzing attentiveness of students in an online classroom environment with emotion detection. In *Innovations in Computational Intelligence and Computer Vision* (pp. 253-261). Springer, Singapore.

Li, S., Deng, W., & Du, J. (2017). Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2852-2861).

Dimitriadis, Y., Martínez-Maldonado, R., & Wiley, K. (2021). Human-centered design principles for actionable learning analytics. *Research on E-learning and ICT in education: Technological, pedagogical and instructional perspectives*, 277-296.

Martinez-Maldonado, R. (2023). Human-centred learning analytics: Four challenges in realising the potential. *Learning Letters*.

Viberg, O., Jivet, I., & Scheffel, M. (2023). Designing culturally aware learning analytics: A value sensitive perspective. In *Practicable Learning Analytics* (pp. 177-192). Cham: Springer International Publishing.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. Psychological review, 110(1), 145.

Martinez-Maldonado, R., Elliott, D., Axisa, C., Power, T., Echeverria, V., & Buckingham Shum, S. (2022). Designing translucent learning analytics with teachers: An elicitation process. Interactive Learning Environments, 30(6), 1077-1091.

D'Mello, S., & Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. Cognition & Emotion, 25(7), 1299-1308.

Lyons, M., Kamachi, M., & Gyoba, J. (1998). The Japanese female facial expression (JAFFE) dataset. *The images are provided at no cost for non-commercial scientific research only. If you agree to the conditions listed below, you may request access to download*.

# Integrating Quantitative Ethnography Methods to Support LA in the Age of AI

**Jun Oshima**
Shizuoka University, Japan
joshima@inf.shizuoka.ac.jp

**Kamila Misiejuk**
University of Bergen, Norway
kamila.misiejuk@uib.no

**Rogers Kaliisa**
University of Oslo, Norway
rogers.kaliisa@iped.uio.no

**Jennifer Scianna, Brendan Eagan**
UW Madison, US
jscianna@wisc.edu, beagan@wisc.edu

**Zack Swiecki**
Monash University, Australia
zach.swiecki@monash.edu

**ABSTRACT**: This workshop explores Quantitative Ethnography (QE) as a framework for supporting learning analytics in the age of Artificial Intelligence (AI). In many learning contexts, we increasingly have access to rich process data. To make meaning of this evidence, our goal is to develop a qualitatively "thick" description of the data and, thus, of learning. However, the more data we have, the more difficult this process becomes: qualitative analysis becomes less feasible, and quantitative analysis becomes less reliable. QE addresses this problem by using statistical techniques to warrant claims about the quality of thick descriptions. The result is a more unified mixed-methods approach that uniquely links the evidence we collect to learning processes and outcomes. This workshop focuses on different quantitative ethnography techniques that address this challenge, including *Epistemic Network Analysis (ENA)* and *Knowledge Building Discourse Explorer (KBDEX)*. The aim of the workshop is to examine these techniques and show how they can be combined to generate a more unified methodology for modeling learning processes and providing actionable insights for research and teaching practices. In addition to showcasing different analysis methods, this workshop includes a presentation of different data coding techniques, including qualitative, AI-supported, and other machine learning methods.

**Keywords**: Quantitative Ethnography, Epistemic Network Analysis, Mixed Methods, Artificial Intelligence, Knowledge Building Discourse Explorer

## 1 BACKGROUND AND PURPOSE

Quantitative Ethnography (QE) seeks to meaningfully analyse and interpret large amounts of rich qualitative data (Eagan, Misfeldt, & Siebert-Evenstone, 2019). Quantitative ethnographic approaches have been used in various fields, including learning analytics, to understand human behaviour and interaction. QE views data documenting learning processes as evidence about the discourse of particular learning cultures (Shaffer, 2017). To make meaning from this evidence and thus gain some understanding of learning processes and outcomes, we must strive for what Geertz (1973) called a qualitatively "thick" description of the data. However, the more data that is available, the more difficult this process becomes: qualitative analysis conducted by hand using traditional methods becomes less feasible; at the same time, quantitative analysis becomes problematic because traditional techniques find large numbers of significant results, some with little theoretical grounding and others with very small effect sizes. QE addresses this problem by using statistical techniques to warrant claims about the quality of thick descriptions. The result is a unified mixed-methods approach that uniquely links the evidence we collect to learning processes and outcomes. QE approach is also useful to ground the learning analytics research in theory by guiding the research and its underlying assumptions, validating models of learning and interpreting the findings (Gašević et al., 2016; Wiley et al., 2020; Rogers et al., 2016).

The main purpose of this workshop is to explore two different network analytic techniques that integrate qualitative and quantitative discourse analysis (Bruun et al., 2017). 1) *Epistemic Network Analysis (ENA)* is a QE technique that models learning processes by constructing networks that represent the cognitive connections learners make in a domain. By modelling patterns of connections in discourse, ENA can help researchers quantify and visualise learning over time for individuals and groups, compare learning across learners or contexts, create trajectories of learning, and model the contributions of individuals to group discourse (Shaffer et al., 2016). 2) Knowledge Building Discourse Explorer (KBDeX) is a QE technique used to analyse and visualise network structures of discourse based on the bipartite graph of words and discourse units. KBDeX can visualise discourse into three different network structures: (1) students, (2) discourse units, and (3) selected words (Oshima et al., 2012). Studies have started to examine learning practices in various contexts from both analytical perspectives (e.g., Oshima et al., 2020).

In addition, this workshop will address the important steps of qualitative data preprocessing, coding, and closing of the interpretative loop. These steps are significant and tightly connected to the theoretical grounding of learning analytics research (Munk et al., 2017), yet rarely considered and discussed. Leveraging the power of AI and traditional qualitative and automated methods, this workshop will showcase the potential and limitations of QE approaches in analyzing text and interaction data. Finally, this workshop will introduce the participants to the concept of *closing the interpretative loop,* which refers to going back to the data to qualitatively validate the results of the quantitative analysis (Arastoopour Irgens & Eagan, 2022).

## 2 INTENDED OUTCOMES, STRUCTURE, AND ORGANIZATION

The workshop is organised both as a mini-conference and hands-on workshop where the participants (a) will be introduced to the QE process, (b) get an overview of a variety of data coding techniques, including qualitative, automated and AI methods, (c) learn about and engage with two QE techniques commonly used in learning analytics research in order to compare the different network approaches, and (d) discuss in small groups how the same data could be analysed with different tools and strategies. These activities will be grounded in QE and will inform a discussion of the philosophical and methodological foundations for network analysis in learning analytics.

This workshop is aimed at participants who are new to QE approach, as well as participants on the intermediate level who would like to deepen their knowledge. Participants from any discipline backgrounds and prior knowledge levels interested in integrating qualitative and quantitative methods in their research can benefit from this workshop.

**During the full-day workshop:**

1. All participants will learn about the QE methodology and foundations.

2. All participants will explore a variety of approaches to data coding and data preprocessing, including AI-supported methods, to prepare different kinds of data for further analysis.

3. All participants will engage in a hands-on workshop introducing them to the basic principles and applications of ENA webtool, rENA, and KBDeX.

4. All participants will be assigned to small groups based on their expressed interest in exploring alternate analytic strategies and discussing the grounding of network approaches to learning analytics in QE.

5. At the end of the workshop, participants will present the main points of the discussions in small groups, which will form the basis for a white paper on QE and learning analytics.

## REFERENCES


Eagan, B., Misfeldt, M., & Siebert-Evenstone, A. (Eds.). (2019). *Advances in Quantitative Ethnography: First International Conference, ICQE 2019, Madison, WI, USA, October 20–22, 2019, Proceedings* (Vol. 1112). Springer Nature.

Arastoopour Irgens, G., & Eagan, B. (2022). The Foundations and Fundamentals of Quantitative Ethnography. In *International Conference on Quantitative Ethnography* (pp. 3-16). Cham: Springer Nature Switzerland.

Bruun, J., Lindahl, M., & Linder, C. (2019). Network analysis and qualitative discourse analysis of a classroom group discussion. *International Journal of Research & Method in Education*, *42*(3), 317-339.

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.

Geertz, C. (2008). Thick description: Toward an interpretive theory of culture. In *The Cultural Geography Reader* (pp. 41-51). Routledge.

Munk, M., Drlík, M., Benko, L. U., & Reichel, J. (2017). Quantitative and qualitative evaluation of sequence patterns found by application of different educational data preprocessing techniques. *IEEE Access*, *5*, 8989-9004.

Oshima, J., Oshima, R., & Matsuzawa, Y. (2012). Knowledge Building Discourse Explorer: a social network analysis application for knowledge building discourse. *Educational Technology Research and Development*, *60*, 903-921.

Oshima, J., Oshima, R., & Saruwatari, S. (2020). Analysis of students' ideas and conceptual artifacts in knowledge-building discourse. *British Journal of Educational Technology*, *51*(4), 1308-1321.

Rogers, T., Gašević, D., & Dawson, S. (2016). Learning analytics and the imperative for theory driven research. *The SAGE Handbook of E-Learning Research,* 232-250.

Shaffer, D. W. (2017). *Quantitative Ethnography.* Madison, WI: Cathcart Press.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, *3*(3), 9-45.

Wiley, K. J., Dimitriadis, Y., Bradford, A., & Linn, M. C. (2020). From theory to action: Developing and evaluating learning analytics for learning design. In *Proceedings of the 10th International Conference on Learning Analytics & Knowledge* (pp. 569-578).

# The 6th Workshop on Predicting Performance Based on the Analysis of Reading and Learning Behavior

**Brendan Flanagan[1*], Atsushi Shimada[2], Fumiya Okubo[2], Hsiao-Ting Tseng[3], Albert C.M. Yang[4], Owen H.T. Lu[5], Hiroaki Ogata[1]**
[1] Kyoto University, Japan
[2] Kyushu University, Japan
[3] National Central University, Taiwan
[4] National Chung-Hsing University, Taiwan
[5] National Chengchi University, Taiwan
[*] flanagan.brendanjohn.4n@kyoto-u.ac.jp

**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 LAK21, LAK22 and LAK23 on secondary school reading behavior and pre/post COVID-19 pandemic changes. Participants of this year's workshop will be given the opportunity to analyze several different datasets, including secondary school prediction of academic performance for more than one subject. 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. In addition, this workshop will accept a wide range of reasearch topics on learning analytics, educational technology, and learning support systems in the post COVID-19 era, including applications of AI in education, proposals for new educational systems, new evaluation methods, and so on.

**Keywords**: Student Performance Prediction, Data Challenge, Reading Behavior, Learning Analytics, Educational Technology

## 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; Flanagan et al., 2021; Flanagan et al., 2022, Flanagan et al., 2023), with 16, 14, 17, 12, 23 and 26 participants respectively.

In addition, challenges from previous years have been updated to include the prediction of academic performance in more than one secondary school subject based on the analysis of reading behavior. Some of the datasets will be offered in a format that is compatible with the OpenLA library (Murata et al., 2020) which can be used by participants to easily implement many common tasks for reading behavior analysis. In the proposed workshop, we will offer a unique opportunity for participants to:

- Analyze large-scale reading log data from secondary school and higher education with performance-based labels for model training.
- Investigate preview, in-class, post-class, and online 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 the data in a real-world learning analytics dashboard.

## 2 OBJECTIVES

While research questions from all participants are welcome, 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

In addition, this workshop accepted a wide range of research topics on learning analytics, educational technology, and learning support systems in the post COVID-19 era, including applications of AI in education, proposals for new educational systems, new evaluation methods, and so on.

Discussion during the workshop focused on the opportunity to integrate the results as part of an ongoing open learning analytics tool development project for inclusion as an analysis feature.

## OVERVIEW

This workshop was held in a mini-track style with a focus on presentations from participant-submitted papers that analyze the data provided by the workshop. In line with the theme of the main LAK conference, Learning Analytics in the Age of AI, the topic of generative AI was strong in many of the submissions: personality traits with ChatGPT for tailored study advice (Hsieh & Yang), Generative AI for at-risk prediction (Liu &Lu; Berr et al.), and reading/learning behavior informed LLM-based Chatbots (Woollaston et al.). Similarly, many submissions continued to investigate the important theme of explainable and trustworthy AI from last year's LAK main conference: evaluation of methods of explanation for AI (Li et al.) and consistency of explainable AI explanations (Hsu & Lu). The traditional task of at-risk prediction was also approached from various perspectives, such as integrating both features based on learner behavior and vector representations of the learning materials with which they were interacting (Shuaileng et al.), self-regulated learning strategies, motivation and programming study behaviors (Wang & Hsu), hierarchical clustering to investigate the relationship of debugging and learning performance (Liu & Hsu), and personalization of the navigation of study materials based on student reading behavior analysis (Ma, Chen, & Lu). There was also a submission that proposes a learning analytics framework for the collection and analysis of affect states and feedback through a emotion focused dashboard. The proceedings of the workshop can be found on the following website: https://sites.google.com/view/lak24datachallenge.

## REFERENCES

Arnold, K. E., & Pistilli, M. D. (2012, April). Course signals at Purdue: Using learning analytics to increase student success. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 267-270). ACM.

Flanagan, B., & Ogata, H. (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*.

Flanagan, B., Majumdar, R., Shimada, A., & Ogata, H. (2021). The 3rd Workshop on Predicting Performance Based on the Analysis of Reading Behavior, *Companion Proceedings of the 11th International Conference on Learning Analytics and Knowledge*.

Flanagan, B., Majumdar, R., Shimada, A., & Ogata, H. (2022). The 4th Workshop on Predicting Performance Based on the Analysis of Reading Behavior, *Companion Proceedings of the 12th International Conference on Learning Analytics and Knowledge*.

Flanagan, B., Majumdar, R., Shimada, A., & Ogata, H. (2023). The 5th Workshop on Predicting Performance Based on the Analysis of Reading Behavior, *Companion Proceedings of the 12th International Conference on Learning Analytics and Knowledge*.

Murata, R., Minematsu, T., Shimada, A. (2020). OpenLA: Library for Efficient E-book Log Analysis and Accelerating Learning Analytics. In International Conference on Computer in Education (ICCE 2020), (pp. 301-306).

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.

# Promoting Open Science in Learning Analytics: An Interactive Tutorial on Licensing, Data, and Containers

**Aaron Haim**
Worcester Polytechnic Institute
ahaim@wpi.edu

**Stephen Hutt**
University of Denver
stephen.hutt@du.edu

**Stacy T. Shaw, Neil T. Heffernan**
Worcester Polytechnic Institute
sshaw@wpi.edu, nth@wpi.edu

**ABSTRACT**: Across the past decade, the open science movement has increased its momentum, making research more openly available and reproducible across different environments. In parallel, learning analytics, as a subfield of education technology, has been increasing as well, providing more accurate statistical models and integrations to improve learning. However, there is a discernible gap between the understanding and application of open science practices in learning analytics. In this tutorial, we will expand the knowledge base towards open data and open analysis. First, we will introduce the complexities of intellectual property and licensing within open science. Next, we will provide insights into data sharing methods that preserve the privacy of participants. Finally, we will conclude with an interactive demonstration on sharing research materials reproducibly. We will tailor the content towards the needs and goals of the participants, enabling researchers with the necessary resources and knowledge to implement these concepts effectively and responsibly.

**Keywords**: Open Science, Reproducibility, Licensing

## 1 INTRODUCTION

Open science and robust reproducibility practices are becoming increasingly adopted within numerous scientific disciplines. Within subfields of educational technology, however, the adoption and review of these practices are sparsely implemented, typically due to a lack of time or incentive to do so (Armeni et al., 2021; Nosek, 2022) with some notable exceptions (Cook et al., 2016; García-Holgado et al., 2021; Makel et al., 2019). Authors have numerous concerns and minimal experience in what can be made publicly available, such as datasets and analysis code (Haim et al., 2023). As such, there is a need for accessible resources, providing an understanding of open science practices, how they can be used, and how to mitigate potential issues that may arise at a later date.

This tutorial aims to expand the knowledge base of participants towards two concepts of open science: open data and open analysis. Participants will be guided through multiple stages of the process, including intellectual property and licensing, data sharing methods, and materials sharing best practices. Throughout the tutorial, we will adapt to the needs and goals of participants, addressing concerns and providing resources tailored to them.

## 2 BACKGROUND

At its core, Open Science seeks to make scientific research, data, and dissemination accessible to all, breaking down the barriers of closed-access publications. It is built on the principles of transparency, collaboration, and shared knowledge. The goals of Open Science are to advance the pace of discovery but also foster a more inclusive, equitable, and accountable scientific community.

As with many things, translation from ideals and principles into real-world implementation comes with considerable challenges. For example, open-access publication typically comes with a higher cost for the researcher (in turn damaging goals of equity and accessibility). Similarly, in education research, data sharing often poses challenges. Data are typically collected in partnership with educators, administrators, and students, who authorize the collection of data for a specific study/set of research questions, and often actively prohibit the distribution of data to third parties. Data can be deidentified, but given how intrinsically personal educational data can be, this task can be labor-intensive. Worse, some of the easier forms of deidentification (such as removing all forum post data prior to sharing[1]) lead to data no longer being usable for a wide range of research and development goals.

Sharing data on a by-request basis and carefully crafting data agreements has long been a potential solution, but it is often ineffective. For example, (Wicherts, Borsboom, Kats, & Molenaar, 2006) contacted owners of 249 datasets, only receiving a response from 25.7%. Within education technology, (Haim et al., 2023) contacted the authors of 594 papers, only receiving a response from 37, or 6.2%, of which only 19 responded that their dataset is public or could be requested. Some of the cited reasons were a lack of rights necessary to release the dataset, personally identifiable information was present, or the dataset itself was part of an ongoing study. The task of sharing data requires significant time investment and can be stalled by changes in email addresses or institutions.

Open Education Science (van der Zee & Reich, 2018), a subfield of Open Science, seeks to address problems of transparency and access, specifically in education research, addressing issues of publication bias, lack of access to original published research, and the failure to replicate. The practices proposed by Open Education Science fall into four categories, each related to a phase of educational research: 1) open design, 2) open data, 3) open analysis, and 4) open publication.Of most relevance to the current tutorial are Open Data and Open Analysis. **Open Data** is about ensuring research data and materials are freely available on public platforms, aiding in replication, assessment, and close examination. However, there can be challenges, especially with educational data. There might be initial agreements that prevent the sharing of data or issues related to personal identifiable information (PII) which restrict what can be made public. **Open Analysis** emphasizes that analytical methods should be reproducible. This is commonly achieved by sharing the code used for analyses on platforms like GitHub or preregistration websites. But there is a catch; the code is often of limited value without the associated data. Simply put, without Open Data, achieving Open Analysis can be

[1] https://edx.readthedocs.io/projects/devdata/en/latest/using/package.html

tough. Moreover, there are challenges like "code rot" and "dependency hell" (as highlighted by Boettiger, 2015), where changing libraries can render older code unrunnable.

# 3 TUTORIAL ORGANIZATION

The proposed tutorial will occur over half a day, focusing on introducing some common open science practices and their usage within learning analytics, along with some interactive examples on how to apply the concepts in research. The target audience is researchers, as the practices offer structure and robustness. Based on past tutorials, we anticipate 5-10 participants and will design an interactive session tailored to their experiences and questions. This approach will allow us to present a responsive tutorial and foster additional community around open science topics.

## 3.1 Prior to the Conference

Prior to the conference, we will be compiling and organizing all relevant resources to be published on a dedicated website for easy access both during and after the tutorial. In addition, we will request all registrants to complete a pre-survey (using the participant registration list following the author registration/early registration deadlines). This survey will gather insights about participants' prior experience with the topics and their specific expectations from the tutorial. We will use this data to customize the tutorial and tailor to the needs of participants.

## 3.2 During the Conference

Our tutorial session will be an interactive and responsive session split into three sections. These sections are outlined below:

1. We will begin the tutorial by discussing how Intellectual Property (IP) intersects with the Open Science Framework. We'll tackle any questions or concerns from attendees with a focus on code licensing, guided by the principles from Creative Commons. We will discuss why licensing code is important, strategies to safeguard a researcher's intellectual property, and provide guidelines for both Tech Transfer and University IP protection.

2. In the next segment of our tutorial, we discuss Open Data relative to the needs of participants. We anticipate opening this section by again addressing participant concerns to frame our future discussion. This will include identifying personal, moral, institutional, or legal concerns regarding open data.

   Participants will be introduced to the concept of Data Enclaves. This will cover understanding the primary objectives of sharing data (including identifying the goals of the individual research team), the relationship between Data Enclaves and GDPR/Privacy legislation, and real-world examples of accessing information via these enclaves. Furthermore, we will provide valuable resources on establishing and efficiently using Data Enclaves. We will also discuss how researchers can share anonymized and synthetic datasets, ensuring the identity of participants remains confidential.

   We will close this segment of the tutorial with a general discussion, weighing the advantages and drawbacks of each approach. Throughout this section, we will emphasize that there is not a "one size fits all" solution and that researchers should make choices based on individual goals and requirements.

3. Finally, we will provide instruction towards sharing materials in a reproducible manner, including best practices on storage, documentation, and privacy. This will be demonstrated with an interactive example using development containers via Visual Studio Code (https://code.visualstudio.com/) and Docker (https://www.docker.com/). The specific example used will be tailored based on survey responses.

### 3.3 Following the Conference

After the conference, all additional resources created for the tutorial will be uploaded to the project's homepage for preservation. As this tutorial wants to repeat and expand upon open science and reproducibility at prior tutorials across conferences, an additional project will be created on the OSF website containing components pointing to all previous conferences and resources. A post-survey will be available at the end and after the tutorial to obtain feedback about the presentation for future use. An aggregate of the response will also be made public on the project's homepage. A Discord channel will be created following the tutorial to foster community on these topics.

## REFERENCES

Armeni, K., Brinkman, L., Carlsson, R., Eerland, A., Fijten, R., Fondberg, R., Heininga, V. E, Heunis, S., Koh, W. Q., Masselink, M., Moran, N., Ó Baoill, A., Sarafoglou, A., Schettino, A., Schwamm, H., Sjoerds, Z., Teperek, M., van den Akker, O. R, van't Veer, A., & Zurita-Milla, R. (2021). Towards wide-scale adoption of open science practices: The role of open science communities. *Science and Public Policy*, 48(5), 605-611. https://doi.org/10.1093/scipol/scab039

Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* **533**, 452–454 (2016). https://doi.org/10.1038/533452a

Boettiger, C. (2015). An introduction to Docker for reproducible research. *ACM SIGOPS Operating Systems Review*, 49(1), 71-79. https://doi.org/10.1145/2723872.2723882

Cook, B. G., Lloyd, J. W., Mellor, D., Nosek, B. A., & Therrien, W. J. (2018). Promoting open science to increase the trustworthiness of evidence in special education. *Exceptional Children*, 85(1), 104-118. https://doi.org/10.1177/0741932516637198

García-Holgado, A., García-Peñalvo, F. J., de la Higuera, C., Teixeira, A., Ehlers, U. D., Bruton, J., Nascimbeni, F., Zea, N. P., & Burgos, D. (2020). Promoting Open Education Through Gamification in Higher Education: the OpenGame project. In *Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality*. 399-404. https://doi.org/10.1145/3434780.3436688

Haim, A., Baxter, C., Gyurcsan, R., Shaw, S. T., & Heffernan, N. T. (2023, July). How to Open Science: Analyzing the Open Science Statement Compliance of the Learning@ Scale Conference. In *Proceedings of the Tenth ACM Conference on Learning@ Scale* (pp. 174-182). https://doi.org/10.1145/3573051.3596166

Makel, M. C., Smith, K. N., McBee, M. T., Peters, S. J., & Miller, E. M. (2019). A path to greater credibility: Large-scale collaborative education research. *AERA Open*, 5(4). https://doi.org/10.1177/2332858419891963.

Nosek, B. (2022). Making the Most of the Unconference. Presented at the *Unconference on Open Scholarship Practices in Education Research*. Available at https://osf.io/9k6pd/

van der Zee, T., & Reich, J. (2018). Open Education Science. *AERA Open*, 4(3). https://doi.org/10.1177/2332858418787466

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, 61(7), 726–728. https://doi.org/10.1037/0003-066X.61.7.726

# LAK24 Assess:

## The 4th Workshop on Learning Analytics and Assessment

**Dragan Gašević and Mladen Raković**
Monash University
{dragan.gasevic, mladen.rakovic}@monash.edu

**Blaženka Divjak**
University of Zagreb
bdivjak@foi.hr

**Yoon Jeon Kim**
University of Wisconsin - Madison
yj.kim@wisc.edu

**Naif Aljohani**
King Abudlaziz University
nraljohani@kau.edu.sa

**Abhinava Barthakur**
University of South Australia
abhinava.barthakur@unisa.edu.au

**ABSTRACT**: The first three editions of the Workshop on Learning Analytics and Assessment were successfully organized at LAK21-23 conferences, resulting in multiple post-workshop collaborations and a special issue in a journal. In this workshop, we intend to address some of the key open challenges in learning analytics that are related to use of learning analytics in formative and summative assessment; measurement of learning progression; reliability and validity of data collection and analysis; and assurance of assessment trustworthiness, in particular given the emergence of the generative artificial intelligence (AI) methods. 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 further formation of a community of practice and possible follow-up publications and special issues in journals.

**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 processes, and improve environments in which learning occurs. Promising results in learning analytics have promoted vibrant research and development activities, and attracted much attention from 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., Gardner & Brooks, 2018), analysis of learning strategies and 21st century skills (e.g., Jovanović et al., 2017), adaptive learner support and personalized feedback at scale (e.g., McNamara et al., 2012; Molenaar, Roda, van Boxtel & Sleegers, 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 Prior Accomplishments of LAK Assess

The first three editions of the Workshop on Learning Analytics and Assessment were successfully organized at LAK21-LAK23 conferences. At each workshop, we gathered 20-30 leading scholars from dynamically emerging fields of learning analytics and assessment. Following the very productive interaction among the workshop participants, this initiative has resulted in multiple post-workshop collaborations and a special issue on Learning Analytics and Assessment in the British Journal of Educational Technology (BJET). To take advantage of this momentum and continue productive discussions on this important and emerging research topic, we propose a fourth edition of the workshop.

### 1.3 Objectives

The main objective of this workshop will be to continue promoting research and practice that looks at the intersection of learning analytics and assessment. We will examine approaches that build upon established principles in educational assessment to improve reliability, validity, usefulness of data collection and analysis in learning analytics. In the workshop, we will also look into the ways how learning analytics can contribute to the future developments in assessment for summative and formative purposes. In addition, we will examine practices for the use of learning analytics to assure assessment trustworthiness, with particular attention to the socio-technical nature of potential challenges. The workshop will also be an opportunity to further frame and shape special issues as important products for the connections between LA and assessment.

## 2 ORGANISATIONAL DETAILS

### 2.1 Proposed Full Day Workshop Schedule

Table 1: Proposed schedule.

| Timing | Description | Contributors/ Facilitators |
|---|---|---|
| 5 minutes | Welcome, introductions and plan for today | Organizers |
| 85 minutes | Assessment using multi-modal learning analytics | Organizers and participants |
| | Morning Coffee | |
| 90 minutes | Learning analytics, assessment and different educational needs | Organizers and participants |
| | Lunch | |
| 90 minutes | Generative artificial intelligence and authentic assessment | Organizers and participants |
| | Afternoon Coffee | |
| 60 minutes | Hands-on activity: use generative AI to aid assessment | Participants |
| 30 minutes | Next steps plenary discussion, and close: Gauge interest in further activities around theory and learning analytics e.g. LAK 2025 workshop, LASI 2025 workshop/tutorial, mid-year check in, etc | Organizers |

### 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 30 October 2023, deadline for abstract submissions 4 Dec 2023, and notification of acceptance 8 Jan 2024. We anticipate a registration of up to 30 participants. #LAKAssess hashtag will be used 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 support further development of a community of practice. The outcomes of the event will be housed on the Google Site. A possible follow-up publications and/or research project proposals will be organized.

## 4 WEBSITE STRUCTURE AND CONTENT

The Google website will: 1. support pre-workshop data gathering and planning materials; 2. act as a collection point for materials, group interactions and archive for the workshop; and, 3. support ongoing dissemination and group activities. It is the aim that the workshop is ongoing, in which case the website will be an ongoing hub for year to year activities and building field memory. The structure of the website is based on theory informing the research cycle, at three stages: design, method, interpretation. Each of these stages will be a section of the website. The website will include: About, 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

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.

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.

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.

# Learning Analytics from Virtual Reality (LAVR)

**Martin Hlosta[1*], Ivan Moser[1], Amir Winer[2], Nitza Geri[2], Umesh Ramnarain[3], Christo van der Westhuizen[3]**
[1] Swiss Distance University of Applied Sciences (FFHS), Switzerland
[2] Open University of Israel, Israel
[3] University of Johannesburg, South Africa
[*] martin.hlosta@ffhs.ch

**ABSTRACT**: The use of (immersive) virtual reality (VR) in educational settings is growing. Thanks to rich sensory data that can be collected from VR applications, this presents many opportunities for learning analytics (LA). The workshop aims to establish first conversations and bring together researchers and practitioners working on topics on the intersection of learning analytics and (immersive) virtual reality in educational settings. Overall, it aims to advance research on the potential and challenges of rich sensory data generated from VR for learning purposes. Ultimately, we strive to better understand how LA can improve the future design of educational VR applications. Therefore, we call for contributions on the role of LA in foundational research about the VR infrastructure and its multimodal analytics; VR for asynchronous learning experiences; and VR for synchronous learning and teaching.

**Keywords**: Virtual Reality, Learning Analytics, Multimodal learning analytics, VR Design, Learning Experiences

## 1 BACKGROUND

For a long time, virtual reality with head-mounted displays was something you would most likely only find in a research lab. However, thanks to advances in technology and falling prices (Goswami, 2023), it has now become affordable (the head-mounted display can be purchased for a similar price to a mobile phone), allowing a significant increase in the number of people using VR.

Following the general trends, there is a growing interest in education to explore the possibilities of VR in the classroom (McGrath et al., 2023), especially in STEM education (Kukulska-Hulme et al., 2023). There is also recent evidence that VR environments can have an impact on applied learning domains (Radianti et al., 2020). These settings are most suitable for medical surgeries (Iop et al., 2022), intelligent manufacturing (Lei et al., 2023), virtual tourism (Melo et al., 2022), teaching arts (Cabero-Almenara et al., 2022), and language acquisition (Dhimolea et al., 2022).

The data collected from a variety of sensors from these devices present a rich source of information to be used for learning analytics. Yet, despite the growing interest in VR and its convergence with learning analytics, the number of papers reporting its opportunities for learning analytics is very scarce. A few examples include Santamaría-Bonfil (2020) and Heinemann et al. (2023).

However, it is evident that an increasing number of VR applications, as well as VR experiences integrated with learning analytics, are emerging from technology companies specializing in VR development (Dwivedi et al., 2022). Claims about the benefits of combining VR and learning analytics lack details of standards, best practices, and academic rigor. Such reports are also almost non-existent (Hwang & Chien, 2022).

As Kukulska-Hulme et al. (2023) mention, apart from the potential of VR, the challenges in education include technical and accessibility issues, together with privacy and security concerns. These concerns also apply to learning analytics. The data generated from VR sensors is more complicated than clicks from virtual learning environments (VLE) and poses additional challenges to data engineering to ensure good quality data for the analysis. The richness of the sensory data poses new challenges for privacy. For example, a recent study on 55,000+ users found out that the motion data from the 100 seconds in a game could identify a user with 94.33% accuracy (Nair et al., 2023).

In addition to the lack of rigorous and public studies, the fact that most research is reported by companies raises some additional issues about the unethical use of learning analytics. These include automated decision-making (performance) enabled by massive data collection without critical evaluation of the underlying collected training data used to develop these models (Carter & Egliston, 2023).

The workshop aims to create a Learning Analytics for Virtual Reality (LAVR) forum for bringing together researchers and practitioners working on topics on the intersection of learning analytics and (immersive) virtual reality in educational settings. Overall, the LAVR workshop aims to advance research on the potential and challenges of rich sensory data generated from VR for learning purposes. Ultimately, we strive to better understand how LA can improve the future design of educational VR applications. Therefore, we call for contributions on the role of LA in foundational research about the VR infrastructure and its multimodal analytics; VR for asynchronous learning experiences; and VR for synchronous teaching in the metaverse. Although this workshop is primarily focused on VR, we are encouraging submission of other eXtended Reality (XR) technologies such as Augmented Reality (AR), Mixed Reality (MR), Haptics, Wearables, etc.

**Topics of interest:**

- Objective vs subjective data analysis
- Multiple sensor merging
- Effective visualizing of the data coming from the VR
- Data preparation and challenges of VR for LA
- Student/teacher acceptance and perception of using VR for LA
- Privacy and security concerns of using LA from VR in education
- LA for the design of VR environments and learning experiences
- LA for performance measurement and evaluation of VR learning
- LA for improving inclusion, equity, and diversity in VR learning environments
- LA for supporting individualized learning processes in VR environments
- LA for enabling and enhancing collaborative learning in VR environments
- LA for supporting integration of VR in hybrid learning environments
- Challenges of algorithmic biases and unintended consequences of LA in VR
- Human-centered explainable LA for VR
- LA for empowering instructors in VR learning environments

- Scalability, availability, and shareability aspects of LA for VR

## 2 OBJECTIVES AND OUTCOMES

We intend to bring for the first time together researchers and practitioners to discuss what possibilities and challenges enable VR for learning analytics. We aim to uncover the emerging trends for this research through both the discussion and a planned keynote presentation. Furthermore, we plan to establish links between the VR for education and the Learning Analytics community. The workshop should also encourage passive participants to work on topics related to LA in VR. As one of the limiting factors is the availability of data from VR systems, the discussion will also focus on how and which datasets can be obtained for the analysis, considering the ethical and privacy issues.

The workshop website with information, a program and the accepted papers has been published and is available at: https://hlostam.github.io/lavr-lak24/

## 3 ACCEPTED PAPERS

Four submissions were accepted for presentation in the workshop, each of them reviewed by at least two members from the Program Committee:

- “Towards the automatization of integrating Learning Analytics into Virtual Reality using xAPI” by Sergej Görzen, Birte Heinemann, and Ulrik Schroeder
- “A Learning Analytics Dashboard to Investigate the Influence of Interaction in a VR Learning Application” by Birte Heinemann, Sergej Görzen, Ana Dragoljić, Lars Meiendresch, Marc Troll, and Ulrik Schroeder
- “Approximating eye gaze with head pose in a virtual reality microteaching scenario for pre-service teachers.” by Ivan Moser, Martin Hlosta, Per Bergamin, Umesh Ramnarain, Christo Van Der Westhuizen, Mafor Penn, Noluthando Mdlalose, Koketso Pila, and Ogegbo Ayodele
- “Towards Learning Analytics for Student Evaluation in the Metaversity” by Amir Winer and Nitza Geri

## REFERENCES

Cabero-Almenara, J., Llorente-Cejudo, C., & Martinez-Roig, R. (2022). The use of mixed, augmented and virtual reality in history of art teaching: A case study. *Applied System Innovation*, *5*(3), 44. https://doi.org/10.3390/asi5030044

Carter, M., & Egliston, B. (2023). What are the risks of virtual reality data? Learning analytics, algorithmic bias and a fantasy of perfect data. *New media & society, 25*(3), 485-504. https://doi.org/10.1177/14614448211012794

Dhimolea, T. K., Kaplan-Rakowski, R., & Lin, L. (2022). A systematic review of research on high-immersion virtual reality for language learning. *TechTrends*, *66*(5), 810-824. https://doi.org/10.1007/s11528-022-00717-w

Dwivedi, Y. K., Hughes, L., Baabdullah, A. M., Ribeiro-Navarrete, S., Giannakis, M., Al-Debei, M. M., ... & Wamba, S. F. (2022). Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy.

*International Journal of Information Management*, *66*, 102542. https://doi.org/10.1016/j.ijinfomgt.2022.102542

Goswami, R. (2023, March 4). Meta announces big price cuts for its VR headsets. CNBC. https://www.cnbc.com/2023/03/03/meta-quest-pro-vr-headset-gets-price-cut.html

Heinemann, B., Görzen, S., & Schroeder, U. (2023). Teaching the basics of computer graphics in virtual reality. *Computers & Graphics, 112*, 1-12. https://doi.org/10.1016/j.cag.2023.03.001

Hwang, G. J., & Chien, S. Y. (2022). Definition, roles, and potential research issues of the metaverse in education: An artificial intelligence perspective. *Computers and Education: Artificial Intelligence*, *3*, 100082. https://doi.org/10.1016/j.caeai.2022.100082

Iop, A., El-Hajj, V. G., Gharios, M., de Giorgio, A., Monetti, F. M., Edström, E., ... & Romero, M. (2022). Extended reality in neurosurgical education: A systematic review. *Sensors*, *22*(16), 6067. https://doi.org/10.3390/s22166067

Kukulska-Hulme, A., Bossu, C., Charitonos, K., Coughlan, T., Deacon, A., Deane, N., Ferguson, R., …, & Whitelock, D. (2023). Innovating Pedagogy 2023: Open University Innovation Report 11. Milton Keynes: The Open University. https://prismic-io.s3.amazonaws.com/ou-iet/4acfab6d-4e5c-4bbd-9bda-4f15242652f2_Innovating+Pedagogy+2023.pdf

Lei, Y., Su, Z., He, X., & Cheng, C. (2023). Immersive virtual reality application for intelligent manufacturing: Applications and art design. *Mathematical Biosciences and Engineering*, *20*(3), 4353-4387. https://doi.org/10.3934/mbe.2023202

McGrath, O., Hoffman, C., & Dark, S. (2023). Future Prospects and Considerations for AR and VR in Higher Education Academic Technology. https://ssrn.com/abstract=4431134

Nair, V., Guo, W., Mattern, J., Wang, R., O'Brien, J. F., Rosenberg, L., & Song, D. (2023). Unique identification of 50,000+ virtual reality users from head & hand motion data. *arXiv preprint arXiv:2302.08927*.

Melo, M., Coelho, H., Gonçalves, G., Losada, N., Jorge, F., Teixeira, M. S., & Bessa, M. (2022). Immersive multisensory virtual reality technologies for virtual tourism: A study of the user's sense of presence, satisfaction, emotions, and attitudes. *Multimedia Systems, 28*(3), 1027-1037. https://doi.org/10.1007/s00530-022-00898-7

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, *147*, 103778. https://doi.org/10.1016/j.compedu.2019.103778

Santamaría-Bonfil, G., Ibáñez, M. B., Pérez-Ramírez, M., Arroyo-Figueroa, G., & Martínez-Álvarez, F. (2020). Learning analytics for student modeling in virtual reality training systems: Lineworkers case. *Computers & Education, 151*, 103871. https://doi.org/10.1016/j.compedu.2020.103871

# Design-Based Research Methodology: Going deeper than methods in multimodal learning analytics

**Danielle Hagood, Daniel Spikol, Liv Nøhr**
University of Copenhagen
dh@ind.ku.dk, ds@di.ku.dk, lsnl@ind.ku.dk

**Vanessa Echeverria, Roberto Martinez-Maldonado, Gloria Nieto, Namrata Srivastava**
Monash University
vanessa.echeverria@monash.edu, roberto.martinezmaldonado@monash.edu, GloriaMilena.FernandezNieto@monash.edu, namrata.srivastava@monash.edu

**Azad Arslan**
University of South Australia
arslan.azad@mymail.unisa.edu.au

**Mutlu Curkurova**
University College London
m.cukurova@ucl.ac.uk

**Daniele Di Mitri**
DIPF
d.dimitri@dipf.de

**Andrew Emerson**
Educational Testing Service
aemerson@ets.org

**Michail Giannakos**
Norwegian University of Science and Technology
michailg@ntnu.no

**Xavier Ochoa**
New York University
xavier.ochoa@nyu.edu

**Yuyu Wang**
University of Wisconsin, Madison
Ywang2466@wisc.edu

**ABSTRACT**: Multimodal learning analytics frequently uses design-based research. In this workshop, we closely consider the methodology underpinning design-based research methods and reflect on how methodology shapes multimodal research. This workshop blends collaborative theoretical reflection and practical knowledge sharing about "doing the work" of multimodal research in learning analytics. In this workshop, participants are positioned as collaborators, and the workshop leaders facilitate discussion by highlighting relevant debates in theory, providing summaries of research, and designing resources and activities to structure reflection, debate, and clarify the methodology underpinning our work.

**Keywords**: multimodal learning analytics, design-based research, methodology

## 1 Workshop Background

As the field of learning analytics matures, work is being done to clarify and strengthen methodology in this research area (Bergner et al., 2018; Chen & Poquet, 2022). With this workshop, we are bringing methodological reflection to multimodal learning analytics (MMLA) by digging into the foundations of design-based research. Over the last decade, multimodal learning analytics has gone from an emerging sub-field of learning analytics research to a key part of understanding how learning unfolds between people and is supported by diverse technologies for collecting data and engaging with learners and teachers (Giannakos et al., 2022, Di Mitri et al., 2023). At this point, this sub-field should dig into the methodological underpinning of work to date and future research.

This workshop closely aligns with one of the conference's topics of interest, which is analytical and methodological approaches, including studies that introduce analytical techniques, methods, and tools for modeling student learning. Anyone with an interest in research design and methodology will find this workshop relevant. The methodological discussion is organized around several current themes in MMLA, providing context in which we can critically evaluate both methodology choices and how things work in real projects. These themes include:

1. Design interventions in the real world, considering the choices we make in complex projects.
2. Learning design and pedagogy, considering the importance of learning and recognizing that research isn't a standalone objective.
3. Ethical privacy, considering the impacts we introduce with interventions and tools, especially AI.

The starting point for this workshop is the distinction between methods (i.e., analyses protocols and data types) and methodology. In this viewpoint, methodology includes the claims, argumentation structures, epistemic framings, and paradigms that underpin research and comprise methodology (Mackenzie & Knipe, 2006). Such considerations are important but especially challenging in design-oriented research such as MMLA. This workshop unpacks the design-based approach to research often adopted in MMLA research to consider and critically reflect on what methodology means in multimodal studies. While the methodology of design-based research has been developed in educational research (e.g., Hoadley, 2004; Kelly, 2004) we are adding to this discussion and building connections to the logistic and technical aspects of MMLA research. This builds multimodal workshops in the past that have emphasized diverse fields and data sources (Spikol et al., 2021).

## 2 Workshop Details

### 2.1 Event Type & Structure

We propose a full-day workshop for up to 40 participants. Both newcomers and experts in multi-modal analyses will be able to participate fully. No technical expertise is necessary; however, participants should be interested in methodology and research design in the field of multimodal learning analytics. The workshop will include reflective and hands-on activities through which the participants and workshop leaders develop a deep understanding and position on design-based methodology in MMLA. These activities will also create research design frameworks, reflective mapping, and useful tools for future work.

### 2.2 Schedule and Activities

#### 2.2.1 *Introduction and Activating Debate: 1 hour*

To start, we will set the tone for active participation and that this is not a "sit and listen" event. Participants will introduce themselves and share their research beliefs and methods knowledge through introducing activities. Departing from this, we will take a round of elevator introductions, providing a more engaging, fun (hopefully), and prime the discussions throughout the day. As facilitators, we will present key definitions for ideas and curated selections from research throughout these activities to frame the activities, inform, and promote debate.

### 2.2.2 *Digging into Design-Based Research: 2 hours*

Different methodological aspects of design-based research are presented, and participants work hands-on with these methods. This section of the workshop will include both short presentations from the organizers (to share knowledge with the participants) and working with this information and the reflection tools in small groups (to share perspectives between participants). This section of the workshop aims to tackle deep issues in design-based research—which may not have definitive answers—but to build a shared perspective, identify differences, and represent these perspectives more systematically.

### 2.2.3 *Advancing methodology in MMLA: 2 hours*

After lunch, the workshop will focus on building bridges between what the participants create in the first half and the technical and logistics of conducting MMLA. An interdisciplinary panel of researchers conducting MMLA will reflect on and give feedback on the theoretical reflections from the morning. The brief for the panel includes briefly presenting their work, with an emphasis on the research design choices and challenges. If participants have submitted a paper to the workshop, they will also be included in this panel. The structure of this phase of the workshop depends on the number of participants and will either be all together or taken in smaller groups so presenters can have around 10 minutes to present.

### 2.2.4 *Reflections and Next Steps: 1 hour*

To conclude the workshop, we want to summarize the outputs and developments. The participants will take a brief reflection survey that allows us to represent our takeaways (text analysis and plots) visually. This will be the starting point for a final discussion. We will also invite participants to join the workshop organizers in turning the workshop outputs into a journal article and briefly outline the publication plan and how they can be involved.

## 2.3 **Recruitment and Dissemination**

This event will be promoted through the CrossMMLA SIG mailing list and the crosmmla.org[1] webpage. This workshop will hold special interest for anyone interested in design-based research or methodology in general, so we plan to partner with related SIGs and research organizations to disseminate this event widely and beyond the learning analytics community.

To facilitate attendance for all, we recognize that some researchers only receive funding to attend conferences if they present a paper. As such, we are also issuing a call for papers and (as described above) have designed a section of the workshop to engage with participants' own work. Thus, the paper call provides strategic inclusive access and will be a meaningful part of the workshop. Paper submissions will be handled through an EasyChair website.

## 2.4 **Equipment**

No special equipment will be needed beyond audio and visual presentation equipment. If more than 20 participants are attending, having a space that can be divided into two rooms is ideal to facilitate break-out activities.

[1] https://crossmmla.org/

## 3 INTENDED OUTCOMES

First, for participants, the outcomes for this workshop include developing or clarifying their own methodological stance towards MMLA based on design-based research. Participants will organize their methodology beliefs and identify how these influence research at different stages and tasks. They will also work with and receive design prompts and organizers that they can use in the future.

Second, for the community at large this workshop provides a venue to assemble careful thought and reflection on methodology in multimodal research. We see the discussions and work from participants as a multidisciplinary panel and wish to consolidate and share the work from this workshop in a paper for the broader learning analytics and education research audience. This paper as well as the participants experiences contributes to advancing methodology in multimodal, design, and learning analytics research.

## REFERENCES

Bergner, Y., Gray, G., & Lang, C. (2018). What does methodology mean for learning analytics?. *Journal of Learning Analytics*, *5*(2), 1–8.

Chen, B., & Poquet, O. (2022). Networks in learning analytics: Where theory, methodology, and practice intersect. *Journal of Learning Analytics*, *9*(1), 1–12.

Di Mitri, D., Schneider, J., & Drachsler, H. (2023). The rise of multimodal tutors in education: Insights from recent research. In O. Zawacki-Richter & I. Jung (Eds.), *Handbook of Open, Distance and Digital Education* (pp. 1037–1056). Springer

Giannakos, M., Spikol, D., Di Mitri, D., Sharma, K., Ochoa, X., & Hammad, R. (Eds). (2022). Introduction to multimodal learning analytics. In *The Multimodal Learning Analytics Handbook* (pp. 3–28). Springer

Hoadley, C. M. (2004). Methodological alignment in design-based research. *Educational Psychologist, 39*(4), 203–212.

Kelly, A. E. (2004). Design research in education: Yes, but is it methodological? *The Journal of the Learning Sciences*, *13*(1), 115–128.

Mackenzie, N., & Knipe, S. (2006). Research dilemmas: Paradigms, methods and methodology. *Issues in Educational Research*, *16*(2), 193–205.

Spikol, D., Ochoa, X., Worsley, M., Di Mitri, D., Cukurova, M., Martinez-Maldonado, R., & Schneider, J. (2021, April). CROSSMMLA Futures: Collecting and analysing multimodal data across the physical and the virtual. In M. Scheffel, N. Dowell, S. Joksimovic, & G. Siemens (Eds.), *Companion Proceedings of the 11th International Learning Analytics and Knowledge Conference (LAK'21)* (pp.381–385). Society for Learning Analytics Research (SoLAR).

# Generative AI for Learning Analytics (GenAI-LA): Exploring Practical Tools and Methodologies

Lixiang Yan, Andy Nguyen, Lele Sha, Jionghao Lin, Mutlu Cukurova, Kshitij Sharma, Roberto Martinez-Maldonado, Linxuan Zhao, Yuheng Li, Yueqiao Jin, and Dragan Gašević

**ABSTRACT**: Generative artificial intelligence (GenAI) presents a transformative opportunity to advance the field of learning analytics (LA). Its capabilities extend from automating the analysis of unstructured data and crafting adaptive educational resources to enhancing the presentation of LA outcomes through rich narratives and detailed explanations. This first GenAI-LA workshop is conceived as a catalyst for dialogue and partnership, spotlighting the potential of GenAI in LA. By assembling a diverse group of learning scientists, LA practitioners, software engineers, and AI specialists, we aim to foster a comprehensive exploration and envisioning of GenAI's pivotal role in advancing LA research and practices.

**Keywords**: generative artificial intelligence, learning analytics, educational technologies

## 1 INTRODUCTION

The progress of generative artificial intelligence (GenAI), exemplified by ChatGPT and other tools employing state-of-the-art large language models (LLMs), revealed its transformative role in boosting human productivity and reshaping the landscape of education (Van et al., 2023; Kasneci et al., 2023). A recent systematic scoping review has identified 53 different use cases for LLMs alone in supporting educational tasks (Yan et al., 2023). However, while ChatGPT has garnered attention, they are but a fraction of the burgeoning GenAI ecosystem. Other GenAI tools, like Midjourney and Whisper, are already transforming sectors like creative arts (Chiu, 2023; Vartiainen & Tedre, 2023) and audio transcription (Gris et al., 2023; Rao, 2023). These novel GenAI technologies could play an essential role in realising the potential of learning analytics (LA) and addressing several of its key challenges, specifically the lack of attempts to intervene in the learning environment (Motz et al., 2023). Nevertheless, how these novel technologies can be embedded in the LA cycle (Clow, 2012) and benefit the development of practical LA solutions with GenAI remains largely unknown.

**The aim of the workshop** is to ignite discussions and collaboration around the potential of GenAI in LA by bringing together a subcommunity of LA researchers and practitioners with a range of expertise in learning sciences, software engineering, and artificial intelligence. In doing so, we plan to address questions such as: What are the different GenAI tools that can support the research and development of LA solutions? How can these tools be embedded into the different stages of the LA cycle, specifically from researching theoretical knowledge and developing prototype products to implementing practical solutions and evaluating intervention effectiveness? What are the opportunities of GenAI in supporting both self-regulated and collaborative learning? Outcomes of this workshop include 1) A consolidated network of LA researchers and practitioners interested in GenAI-LA; 2) A workshop proceeding featuring pioneering works integrating GenAI within LA research and practices.; and 3) An open-source toolkit geared towards embedding GenAI in LA projects.

## 2 BACKGROUNDS

The surge in research interest in GenAI was catalysed by the public unveiling of ChatGPT in November 2022. Following this, numerous research initiatives have been undertaken to explore the applications and implications of these emerging technologies in the educational sector (Kasneci et al., 2023). Studies focused on ChatGPT and other LLM-based tools have highlighted their capabilities in providing comprehensive feedback that articulates students' performance more effectively than human instructors (Dai et al., 2023). Furthermore, these models have demonstrated superior performance in reflective writing assessments compared to average students (Li et al., 2023) and have enhanced engagement in digital formative assessments by facilitating conversation-based assessments (Yildirim-Erbasli & Bulut, 2023). Beyond LLMs, text-to-image GenAI models, such as DALL-E 2 and Midjourney, have paved their way into the creation of teaching and learning materials, supporting visual learning in domains like medical training (Mazzoli, Semeraro, & Gamberini, 2023) and craft education (Vartiainen & Tedre, 2023). Speech-to-text models, exemplified by OpenAI's Whisper, have been utilized for tasks like transcription of educational videos (Rao, 2023) and documentation of collaborative discourse between learners (Cao et al., 2023). The advent of other GenAI models, including text-to-code (Advanced Data Analysis) and text-to-audio (Voicebox), has expanded the repertoire of tools available for LA researchers and practitioners in data analysis and stakeholder communication. The introduction of large multimodal models, such as GPT-4, has further broadened the scope and applicability of GenAI in diverse educational contexts. For instance, GPT-4's ability to understand and generate content across multiple modalities has been instrumental in creating adaptive learning environments where textual, visual, and auditory information can be seamlessly integrated, offering learners a more holistic and immersive educational experience. The potential of these GenAI technologies may be pivotal in translating insights from prior LA research into tangible solutions for daily educational practices, potentially addressing the observed disconnect between LA's objectives and its academic contributions (Motz et al., 2023). Nonetheless, the challenges associated with GenAI, particularly concerning its trustworthiness and transparency, warrant consideration, given the existing ethical complexities in LA (Tsai et al., 2020).

### 2.1 Evidence of interest

Recent conferences and seminars, such as the 24th International Conference on Artificial Intelligence in Education (AIED 2023) and the Eighteenth European Conference on Technology Enhanced Learning (ECTEL 2023), have observed an uptick in discussions related to GenAI. A substantial number of papers from these events have emphasized the application of GenAI, especially LLMs, in educational settings. This proliferation of GenAI-related publications indicates the growing interest and exploration of this technology within the educational research domain. However, its integration within the LA community remains in the early stages. This workshop intends to foster this emerging community, offering a dedicated space for exploring the potential and challenges of GenAI-LA.

## 3 ORGANISATIONAL DETAILS

### 3.1 Workshop format, participation, and pre-workshop task

The workshop is scheduled as a half-day, in-person event, accommodating between 15 to 30 participants. The anticipated attendees encompass a spectrum ranging from learning scientists and LA practitioners to software engineers and AI specialists, all converging on the application of GenAI in LA research and practices. The workshop is open to all interested parties, regardless of their proficiency level in the field. Participants are encouraged to present prototype concepts or initial projects pertaining to GenAI-LA for deliberation and collaborative activity during the workshop. Upon approval of this workshop proposal, a call for papers will be disseminated to invite more elaborate contributions to this area. Submissions for the workshop, between 2-4 pages, will undergo a review process led by the organising committee and paper authors. Workshop attendees will have access to both submitted and approved papers in advance, facilitating informed discussions during the event. Furthermore, participants will be required to complete a pre-workshop survey. This survey aims to gather data on participants' prior engagements with GenAI and prospective ideas for its application in research. Such information will serve as a foundation for fostering discussions and collaborations during the workshop.

### 3.2 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 (March 18 or 19, 2024). The workshop has five parts:

1. **Overview of GenAI-LA (60 mins)**. In the first part of the workshop, and based on the survey results, we will present an overview of the state-of-the-art GenAI and their potentials in LA, including technology demonstrations, system architectures, off-the-shelf applications, focusing on discussing both the opportunities and challenges of adopting GenAI in LA research and practices.

2. **Prototype Showcases (60 mins)**. The second part will be for authors of the accepted workshop papers to provide a brief overview of their works as flash presentations. They will be able to prepare 6 slides to be presented in 20 seconds each so each will provide a brief 2-minutes presentation. An invited discussant with expertise in both LA and GenAI will provide feedback on each presentation and spark discussion among the audiences for the next part.

3. **Collaborative Design Sessions (90 mins)**. The third part will be a group-based activity. Participants will be divided into small groups based on their experiences and interests. We will ensure a mixture of expertise and experience with GenAI in each team with a shared common interest. Each team will choose one of the presented prototypes and work together to refine, enhance, or brainstorm around the presented prototypes (60 mins). To facilitate structured discussion, participants will be asked to use the SWOT analysis framework to identify and analyse the prototype's strengths, weaknesses, opportunities, and threats. At the end of the activity, groups will present the results of their SWOT analysis on the particular prototype to the entire workshop (30 mins). During the activity, organisers experienced in GenAI and LA will be available for guidance, ensuring teams are on the right track and assisting with potential challenges.

4. **Discussion on next steps (30 mins)**. All participants will be invited to contribute with ideas to set a potential GenAI-LA research agenda.

### 3.3 Dissemination strategy

Upon approval of this workshop, a dedicated website will be established. The website will serve as the primary platform for disseminating a call for participation. Additionally, outreach will be conducted via Twitter accounts and mailing lists accessible to the workshop organisers. The website will feature essential information, including the workshop's objectives, details about the organisers, contact information, and subsequent reports and outputs from the event. Accepted submissions will be made available either within the LAK companion proceedings, or as part of a CEUR proceeding.

### 3.4 Logistics and tools

The workshop is planned as an in-person event. The selected venue will feature adaptable seating arrangements, with movable desks and chairs, catering for the collaborative design session. For pre-workshop interactions, a Google form will be utilised to distribute the pre-workshop survey. This distribution will also contain an invitation to a dedicated Slack channel, ensuring seamless communication both before and after the workshop. Attendees are advised to bring personal computers or laptops to engage with various GenAI prototypes during the session.

## REFERENCES


Cao, J., Ganesh, A., Cai, J., Southwell, R., Perkoff, E. M., Regan, M., ... & D'Mello, S. (2023, June). A Comparative Analysis of Automatic Speech Recognition Errors in Small Group Classroom Discourse. In *Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization* (pp. 250-262).

Chiu, T. K. (2023). The impact of Generative AI (GenAI) on practices, policies and research direction in education: a case of ChatGPT and Midjourney. *Interactive Learning Environments*, 1-17.

Clow, D. (2012, April). The learning analytics cycle: closing the loop effectively. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 134-138).

Dai, W., Lin, J., Jin, F., Li, T., Tsai, Y. S., Gasevic, D., & Chen, G. (2023). Can large language models provide feedback to students? A case study on ChatGPT.

Gris, L. R. S., Marcacini, R., Junior, A. C., Casanova, E., Soares, A., & Aluísio, S. M. (2023). Evaluating OpenAI's Whisper ASR for Punctuation Prediction and Topic Modeling of life histories of the Museum of the Person. *arXiv preprint arXiv:2305.14580.*

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, *103*, 102274.

Li, Y., Sha, L., Yan, L., Lin, J., Raković, M., Galbraith, K., ... & Chen, G. (2023). Can large language models write reflectively. *Computers and Education: Artificial Intelligence, 4*, 100140.

Mazzoli, C. A., Semeraro, F., & Gamberini, L. (2023). Enhancing Cardiac Arrest Education: Exploring the potential use of MidJourney. *Resuscitation, 189*.

Motz, B. A., Bergner, Y., Brooks, C. A., Gladden, A., Gray, G., Lang, C., ... & Quick, J. D. (2023). A LAK of direction: Misalignment between the goals of learning analytics and its research scholarship. *Journal of Learning Analytics*, 1-13.

Rao, A. (2023). Transcribing Educational Videos Using Whisper: A preliminary study on using AI for transcribing educational videos. *arXiv preprint arXiv:2307.03200.*

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).

Van Dis, E. A., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: five priorities for research. *Nature, 614*(7947), 224-226.

Vartiainen, H., & Tedre, M. (2023). Using artificial intelligence in craft education: crafting with text-to-image generative models. *Digital Creativity, 34*(1), 1-21.

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., ... & Gašević, D. (2023). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, in press.

Yildirim-Erbasli, S. N., & Bulut, O. (2023). Conversation-based assessment: A novel approach to boosting test-taking effort in digital formative assessment. *Computers and Education: Artificial Intelligence, 4*, 100135.

# Workshop in New Measures & Metrics in Education

**Charles Lang[1], Geraldine Gray[2], Christopher Brooks[3], Yoav Bergner[4]**
[1] Teachers College Columbia University, New York, USA
[2] Technology University, Dublin, Ireland
[3] University of Michigan, Ann Arbor, USA
4 Yoav. Bergner, New York University, New York, USA
charles.lang@tc.columbia.edu, Geraldine.Gray@tudublin.ie,
brooksch@umich.edu, yoav.bergner@nyu.edu

Learning analytics offers tremendous potential to improve educational outcomes, but new measures and metrics often remain isolated within institutions or companies. This half-day workshop at the 2024 Learning Analytics and Knowledge Conference aims to collaboratively advance the development and dissemination of innovative analytic measures in education. Submissions of new and innovative metrics will be compiled on a website and presented at the event. Through mini-presentations, structured discussion, and breakout sessions, participants will exchange insights about creating, validating, and distributing novel metrics. The workshop will conclude by voting on the most promising measure and awarding a prize. By synthesizing diverse viewpoints, the workshop intends to catalyze the evolution and adoption of impactful new techniques in learning analytics. Outcomes will be shared through a public website, potential publications, and continued online community dialogue. This interactive workshop provides an exciting opportunity to collectively spur progress in developing the next generation of learning metrics.

**Keywords**: Methods, measures, metrics

## 1 ORGANIZATIONAL DETAILS OF PROPOSED EVENT

### 1.1 Motivation

Currently, new analytic measures remain siloed within institutions and companies [1,2,3,4]. As a result, they lack the testing and refinement that comes with broader exposure, debate and input [7]. This workshop will bring together researchers and practitioners from the classroom, industry and policy to facilitate collaborative ideation, refinement and dissemination of new measures and metrics in learning analytics. The ultimate goal is to help develop pipelines for innovative and useful new measures from wherever they originate, be it the university, the school, or the product development team, to wider use.

The adoption of novel learning analytics measures face substantial barriers within educational institutions. Many schools lack the financial resources, staff, infrastructure, and technical capabilities needed to implement new measures [10]. Without evidence demonstrating validity and impact, institutions are often reluctant to devote limited resources to unproven metrics that may not integrate well with existing data systems [7]. While privacy and ethical concerns surrounding data use further complicate adoption [12]. Additionally, new metrics may misalign with established assessments and accreditation standards favored by administrators and faculty who tend to resist altering familiar practices [11]. Given limited budgets, skepticism about unvalidated measures, technical integration challenges, apprehension about data ethics, and organizational inertia, institutions demonstrate understandable caution in adopting innovative learning analytics metrics [6].

The learning analytics community has an interest in addressing these concerns in a methodical and impactful way. One that tackles both the iterative improvement of measures and aids in overcoming barriers to their adoption. One strategic approach is to surface, test and promote new measures – a role this workshop aims to facilitate.

### 1.2 Objectives

The objectives of the proposed workshop are twofold. First, it aims to provide a forum to discuss, debate and advance the development of new measures and metrics in education. Second, the workshop will focus on understanding and overcoming obstacles to developing, validating and disseminating innovative measures and metrics. By exchanging knowledge and experiences, participants can gain insights into challenges and strategies to operationalize and distribute metrics more effectively.

### 1.3 Format

After a general introduction and framing of the problems. The key activity will be the presentation and discussion of specific new and innovative measures that participants have submitted prior to the workshop. Participants will be requested to submit a measure or metric that they have or are currently working on that they believe is novel in some way - Measures do not necessarily need to be newly created but they should be new to the broader educational community. The workshop is methodologically agnostic and submissions of any type of measure, whether qualitatively or quantitatively derived are encouraged. Submissions will include a brief description, a sample data set (real or manufactured) and a visualization submitted through Github. Participants will be encouraged to keep submissions brief to lower the barrier to entry. At the end of the workshop, the most promising measures will be voted on to be award a small prize and assistance from the workshop organizers to make their measure usable by a broader audience. These awards will be the key outcome of the event.

### 1.4 Dissemination

The workshop outcomes will be disseminated through multiple channels. A public website thorough Github sites will compile promising metrics and serve as a reference for the community. The organizers also intend to synthesize insights and produce a review of the measures presented and published as a CEUR submission. During the workshop, participants will be invited to share descriptions, sample data and visualizations for their metrics. These contributions may be published on the workshop website as well. To continue conversations after the event, the organizers will facilitate online community discussions through platforms like GitHub and Slack. The specific mediums will be determined based on participant preferences. Key outcomes and follow-up activities will also be summarized in slides and documents that are openly accessible. Through these multifaceted efforts, the workshop aims to advance the development and availability of impactful new metrics in learning analytics.

## 2 CONTRIBUTING PRESENTATIONS

| **Authors** | **Measure/Metric** |
|---|---|
| Joseph Thibault | Cursive Recorder |
| Marc Aimé Tchoumado | Statistiques |
| Peter Ruijten-Dodoiu | Self-reflection on personal development |
| Laura McReavy Hearnsberger | Engagement Culture |
| Jill-Jênn Vie, Samuel Girard | Reward / conditional average treatment effect |
| Paul Gamper | Solution/error - space of a given programming task |
| Arun Lekshmi Narayanan | MADD++ |
| Harry Lin | MUAS |

## 3 REFERENCES

1. El Alfy, S., Marx Gómez, J. and Dani, A., 2019. Exploring the benefits and challenges of learning analytics in higher education institutions: A systematic literature review. *Information Discovery and Delivery*, *47*(1), pp.25-34.

2. Klasen, D. and Ifenthaler, D., 2019. Implementing learning analytics into existing higher education legacy systems. *Utilizing learning analytics to support study success*, pp.61-72.

3. Wong, B.T.M. and Li, K.C., 2020. A review of learning analytics intervention in higher education (2011–2018). *Journal of Computers in Education*, *7*(1), pp.7-28.

4. Tsai, Y.S., Rates, D., Moreno-Marcos, P.M., Muñoz-Merino, P.J., Jivet, I., Scheffel, M., Drachsler, H., Kloos, C.D. and Gašević, D., 2020. Learning analytics in European higher education—Trends and barriers. *Computers & Education*, *155*, p.103933.

5. Joksimović, S., Kovanović, V., & Dawson, S. (2019). The journey of learning analytics. *HERDSA Review of Higher Education*, *6*, 27-63.

6. Klein, C., Lester, J., Rangwala, H., & Johri, A. (2019). Technological barriers and incentives to learning analytics adoption in higher education: Insights from users. *Journal of Computing in Higher Education*, *31*, 604-625.

7. 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.

8. Rienties, B., Nguyen, Q., Holmes, W., & Reedy, K. (2017). A review of ten years of implementation and research in aligning learning design with learning analytics at the Open University UK. *Interaction Design and Architecture (s)*, *33*, 134-154.

9. Gašević, D., Greiff, S., & Shaffer, D. W. (2022). Towards strengthening links between learning analytics and assessment: Challenges and potentials of a promising new bond. *Computers in Human Behavior*, *134*, 107304.

10. Ifenthaler, D. (2021). Learning analytics for school and system management. *OECD Digital Education Outlook 2021 Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots: Pushing the Frontiers with Artificial Intelligence, Blockchain and Robots*, 161.

11. Kaliisa, Rogers, Anders Kluge, and Anders I. Mørch. "Overcoming challenges to the adoption of learning analytics at the practitioner level: A critical analysis of 18 learning analytics frameworks." *Scandinavian Journal of Educational Research* 66.3 (2022): 367-381.

12. Kitto, K., & Knight, S. (2019). Practical ethics for building learning analytics. *British Journal of Educational Technology*, *50*(6), 2855-2870.

# Data Storytelling Narratives and Learning Analytics

**Gloria Milena Fernandez-Nieto [1*], Vanessa Echeverria[1], Roberto Martinez-Maldonado[1], Yi-shan Tsai [1], Lu Lawrence[2], Shaveen Singh[1], Stanislav Pozdniakov[3], Lujie Karen Chen[4], Jiaqi Gong[5], Louise Yarnall[6]**

Monash University [1], Utah State University [2], University of Queensland [3], University of Maryland Baltimore County [4], The University of Alabama [5], SRI International [6]

[1*] gloriamilena.fernandeznieto@monash.edu

**ABSTRACT**: Data Storytelling (DS) in Learning Analytics (LA) has proven as an effective approach to communicating insights to non-data experts (e.g., students and teachers). DS brings the promise to incorporate narratives into LA interfaces (e.g., dashboards) to facilitate the provision of direct feedback and pedagogical explanations. The LA community has researched Data Storytelling principles and techniques to support educational stakeholders in interpreting their teaching and learning progress. However, given the relevance of the story narrative, challenges arise to provide unbiased, fair, and meaningful stories without misleading the communication of insights. This workshop aims to explore the formal and practical challenges and opportunities of DS by engaging in discussions with the LA community. In this workshop, we expect to spark discussion on these main topics: What methods and methodologies of DS from other domains are suitable for LA? How to evaluate the impact of DS in LA? How can we automate the process of generating fair and unbiased data stories to facilitate sense-making and effectively communicate insights? This workshop will bring together storytelling researchers and practitioners, whose data storytelling in LA is a special case, to clarify and converge on the future of DS in LA related to their challenges and opportunities.

**Keywords**: educational data storytelling, explainable dashboards, visual learning analytics

## 1 Workshop Background

### 1.1 Motivation

There is a growing interest in creating Learning Analytics (LA) interfaces (e.g., dashboards, visualizations, or reports) to support educational stakeholders in monitoring learning tasks (Salas-Pilco et al., 2022). However, recent literature reviews and empirical studies report that most of these LA interfaces have serious limitations, such as showing visualizations that are difficult to understand by non data experts (Corrin & de Barba, 2015; Herodotou et al., 2019), lack of effectiveness in communicating insights (Bodily et al., 2017), and failing to align with educators' pedagogical needs (Kaliisa et al., 2022; Sergis et al., 2017).

Commonly current research and design approaches adopted to create LA interfaces are generating interfaces that are often hard to interpret in a timely manner (Duval, 2011). With the increasing amount of complex data traces captured (in online and physical spaces), there is a need for compelling ways to distill information into meaningful, memorable, and engaging insights (Dominyk, 2022). One of the strategies to address these challenges is the improvement of the *explanatory* *design features* of current LA interfaces. Data storytelling (DS) techniques and principles provide a way to include narrative and elements to explain and connect the learning design goals with visual elements aiming at guiding the user's attention to relevant insights. For instance, Echeverria et al. (2018) demonstrated the potential of enhancing visualisations with DS visual elements (e.g., title, highlights, shaded areas) in helping teachers explore visualisations with less effort. Similarly, Martinez-Maldonado et al. (2020) demonstrated the promise of using a layered storytelling approach to communicate insights on team performance. Following a similar layered approach, Fernandez-Nieto et al. (2021) crafted data stories to promote students' reflections. Their work demonstrated that learner data stories were useful for students to identify potential improvements and errors they performed while enacting clinical simulations. One of the most recent works on DS is presented by Pozdniakov et al. (2023). The authors evaluated the impact of teachers' visualisation literacy on their interactions with LA dashboards, and found that teachers with low visualisation literacy especially benefited from DS-based visual guidance. While these prior works have demonstrated how DS can benefit teachers and students when interpreting data from LA interfaces, there are still challenges and opportunities to explore in terms of DS automation, ethics, fairness, scalability, and impact (Fernandez-Nieto et al., 2021; Martinez-Maldonado et al., 2020; Zdanovic et al., 2022).

This workshop aims to explore formal and practical approaches to actively increase DS adoption and impact in the LA community. Thus, this workshop will provide a scenario for participants to reflect and critically discuss the following aspects: What methods and methodologies of DS from other domains are suitable for LA? How can researchers effectively evaluate the impact of DS in LA? How can we automate the process of generating fair and unbiased data stories? To start, the organisers of the workshop will provide a review of lessons learnt from using DS to support stakeholders' interpretations of visual interfaces from the current literature in LA and other research fields such as Information Visualization (InfoVis) and Human-Computer Interaction (HCI). This review will be a starting point to open the discussion with participants regarding challenges (e.g., automate DS according to particular needs) and opportunities (e.g., the use of AI to generate narratives to support interpretations) of DS to effectively support educational stakeholders to make sense of their data traces and make them actionable to improve their practice.

Topics of interest include data storytelling work, which encompasses case studies, interactive visualisations, and narrative-driven stories. These are detailed as follows:

- Data storytelling for impact: How can data storytelling be used to communicate learning insights and inform students/teachers actions?
- Developing and evaluating methods and methodologies of data storytelling: What are the most effective ways to tell learning/teaching stories with data?
- Evaluating and measuring the impact of data storytelling approaches on learning outcomes: How can we measure the effectiveness of data storytelling in different learning contexts?
- Designing and implementing automated data storytelling tools and techniques: Can we develop tools and techniques to automate the process of creating learning/teaching stories?
- Addressing bias and fairness in data storytelling approaches: How can we ensure that learning/teaching stories are fair and accurate representations of reality?

### 1.2 Objectives

The main objective of this workshop will be to continue promoting research and practice that looks at the intersection of learning analytics and data storytelling. Particularly, the aims of this workshop includes: 1) enable researchers and practitioners to explore the challenges and opportunities that DS may incorporate to their practices when designing LA dashboards, and 2) have a holistic view of formal and practical work that are currently used in the LA community to incorporate DS into their designs and practices. Our workshop included a call for papers for researchers to share their current work on DS in LA, aiming to enable further discussions.

## 2 Workshop Details

### 2.1 Proposed Half-Day Workshop Schedule

The workshop is scheduled for Tuesday, March 19, 2024, and will be held in person from 1:30pm to 5:00pm.

*2.2.1 Welcome and contextualisation of existing work on DS: 45 minutes:* In the first activity of this workshop the organisers will introduce current work on DS in LA and other research fields such as Information Visualisation and Human-Centered Design. After the initial opening presentation, we will encourage active participation in small groups for participants to share their perceptions and motivations in terms of the existing DS research and what challenges and opportunities they foresee.

*2.2.2 Methods and Methodologies for DS in LA Dashboards: 2 hours.* Accepted papers will facilitate a 15-minute presentation. Each presenter will explain their formal or practical DS approaches and indicate the challenges they have faced running their studies.

*2.2.3 Reflections and Roadmap: 45 minutes.* The last part of the workshop will focus on reflecting on the work presented during the session and defining strategies to increase active use of DS in LA research and design practices. As a result of this session, a list of challenges and opportunities will be identified and used to co-create a roadmap to define priorities and main challenges for the DS LA community.

### 2.2 Intended outcomes

The workshop website information, program, and accepted papers has been published and are available at: DS-LAK24 Website. The website will: 1. support pre-workshop data gathering and provide planning materials; 2. facilitate the collection of materials and document the interactions of groups attending the workshop; and 3. aid in the ongoing dissemination of information and support group activities. The goal is for the workshop to be an ongoing event. In this case, the website will serve as a continuous hub for activities year after year, contributing to the building of field memory.

Finally, the accepted papers will be published in the CEUR[1] Workshop Proceedings.

[1] https://ceur-ws.org/

## REFERENCES


Salas-Pilco, S. Z., Xiao, K., & Hu, X. (2022). *Artificial intelligence and learning analytics in teacher education: A systematic review. Education Sciences,* 12(8), 569.

Corrin, L. & De Barba, P. Exploring students' interpretation of feed- back delivered through learning analytics dashboards, in *Proc. 31st Annu. Conf. Australas. Soc. Comput. Learn. Tertiary Educ.*, 2014, pp. 629–633. [Online]. Available: http://www.ascilite.org/conferences/ dunedin2014/proceedings/

Kaliisa, R., Mørch, A.I. & Kluge, A. *My Point of Departure for Analytics is Extreme Skepticism': Implications Derived from An Investigation of University Teachers. Learning Analytics Perspectives and Design Practices*. *Tech Know Learn* 27, 505–527 (2022). https://doi.org/10.1007/s10758-020-09488-w

Sergis, S., Sampson, D.G. (2017). Teaching and Learning Analytics to Support Teacher Inquiry: A Systematic Literature Review. In: Peña-Ayala, A. (eds) *Learning Analytics: Fundaments, Applications, and Trends. Studies in Systems, Decision and Control*, vol 94. Springer, Cham. https://doi.org/10.1007/978-3-319-52977-6_2

Herodotou, Christothea & Rienties, Bart & Boroowa, Avinash & Zdráhal, Zdenek & Hlosta, Martin. (2019). A large-scale implementation of predictive learning analytics in higher education: the teachers' role and perspective. *Educational Technology Research and Development*. 67. https://doi.org/10.1007/s11423-019-09685-0.

Bodily, R., & Verbert, K. (2017, March). Trends and issues in student-facing learning analytics reporting systems research. In *Proceedings of the seventh international learning analytics & knowledge conference* (pp. 309-318).

Dominyk Zdanovic, Tanja Julie Lembcke, and Toine Bogers. 2022. The Influence of Data Storytelling on the Ability to Recall Information. In *Proceedings of the 2022 Conference on Human Information Interaction and Retrieval (CHIIR '22)*. Association for Computing Machinery, New York, NY, USA, 67–77. https://doi.org/10.1145/3498366.3505755

Echeverria, V., Martinez-Maldonado, R., Buckingham Shum, S., Chiluiza, K., Granda, R., & Conati, C. (2018). Exploratory versus Explanatory Visual Learning Analytics: Driving Teachers' Attention through Educational Data Storytelling. *Journal of Learning Analytics*, *5*(3), 73—97. https://doi.org/10.18608/jla.2018.53.6

Martinez-Maldonado, R., Echeverria V., Fernandez Nieto G., & Buckingham Shum S. 2020. From Data to Insights: A Layered Storytelling Approach for Multimodal Learning Analytics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376148

Pozdniakov S., Martinez-Maldonado R., Tsai Yi-Shan, Echeverria Vanessa, Namrata Srivastava, & Gasevic Dragan. 2023. How Do Teachers Use Dashboards Enhanced with Data Storytelling Elements According to their Data Visualisation Literacy Skills? In *LAK23: 13th International Learning Analytics and Knowledge Conference (LAK2023)*. Association for Computing Machinery, New York, NY, USA, 89–99. https://doi.org/10.1145/3576050.3576063

Fernandez-Nieto, G. M., Echeverria, V., Shum, S. B., Mangaroska, K., Kitto, K., Palominos, E., ... & Martinez-Maldonado, R. (2021). Storytelling with learner data: Guiding student reflection on multimodal team data. *IEEE Transactions on Learning Technologies*, 14(5), 695-708. https://doi.org/10.1109/TLT.2021.3131842.

# Advancing Learning Analytics with Complex Dynamical Systems: Trends and Challenges in Non-Linear Modeling of Learning Data

**Oleksandra Poquet**
**Technical University of Munich**
sasha.poquet@tum.de

**Muhterem Dindar**
Tampere University
muhterem.dindar@tuni.fi

**Laura Allen**
University of Minnesota
lallen@umn.edu

**Elizabeth Cloude**
Tampere University
elizabeth.cloude@tuni.fi

**Daryn Dever**
University of Central Florida
daryn.dever@ucf.edu

**ABSTRACT**: Although there is a complex interplay between cognitive, motivational, social, and affective processes during learning, current Learning Analytics (LA) frameworks often overlook the dynamics of these processes. Existing analytical and computational methods are ill-equipped to address these complexities. Thinking and methods in complex dynamic systems (CDS) hold significant potential for addressing these challenges, however, their integration in LA remains limited. This international workshop addresses this gap. This workshop aims to both educate the broader LA community about the potential of CDS, as well as help researchers who are currently applying these methods to learning data to identify common challenges in their work and transform the status quo. The participants will explore CDS applications in various learning analytics areas, such as in writing, self-regulated learning, emotion regulation, and social learning, and in a variety of settings, including game-based environments, intelligent tutoring systems, computer-assisted learning, among others. The participants will both have hands-on experience with selected methods and exposure to the current LA applications of CDS.

**Keywords**: learning analytics, complexity science, common challenges

## 1 WORKSHOP ORGANIZERS

The workshop will be organized by five learning analytics researchers: three from European institutions, and two from the US. All the organizers are active in applying complex dynamical systems (CDS) perspectives in their learning analytics work. They bridge different scholarly groups within the Society and apply CDS approaches to areas such as writing, social learning, self-regulation, and emotion regulation. The workshop will bring together researchers working across these various thematic areas of learning analytics who have interest in advancing conceptual and analytical tools they currently use. The organizers will use their scholarly networks to engage invited speakers outside of learning analytics who will offer expert input from the CDS perspective.

## 2 WORKSHOP BACKGROUND

Learning is a nonlinear and cyclical process where learners rely on various cognitive, metacognitive, motivational, and affective processes that continuously interact within a learning context. However, state-of-the-art conceptual and analytical frameworks in the Learning Analytics (LA) field are limited in explaining how these dynamic processes occur. In LA is that learner behaviors are often profiled by ignoring the temporality and critical fluctuations in students' behavior (Poquet et al., 2023).

Reductionist analytical methods decompose a whole into its parts, often losing information about the underlying dynamic, multi-level processes of how these parts interact. These parts are then examined separately, often under an assumption that their effects are additive (i.e., component-dominant). However, this is rarely the case in systems with changing components and interactions such as learning and educational systems that involve interdependent components that can interact in non-additive ways (Jacobson et al., 2016). Computational analyses are in dire need of the interaction-dominant methods - those that can describe upstream effects of the micro-level processes and the downstream effect of the macro-level processes, in a non-additive manner (Richardson, Dale, & Marsh,2014).

Numerous scholars have called out these methodological limitations in educational and psychological quantitative research, claiming that current methodologies are insufficient for capturing the complexity inherent in deep learning processes (Hilpert & Manchard, 2018). Much of the critique has come from recurrent arguments by the proponents of complex dynamical systems (CDS) as a framework for educational research. Several themes are common to these critiques:

- Data aggregation (e.g., averaging) does not adequately describe change in each learner. Popular statistical methods in psychological and educational research privilege such data aggregation. However, learning processes (cognition, affect, etc.) follow person-specific dynamic models (Molenaar & Campbell (2009) where measures of a singular learner can continuously vary.
- Conventional statistics based on the central limit theorem assume that data distribution has certain properties (e.g., normal distribution, regression to the mean, etc.). Yet, data collected from learning environments often follow non-normal, heteroscedastic, and non-linear distributions, suggesting interdependencies and complexity within the data.
- Statistical models for time series often require data transformation where intra-individual variability and event history are removed. They also embed assumptions of some randomness and independence of observations. Yet, learning events may be contingent on one another.
- Methods like regression analysis and similar apply an additive logic in estimating factors influencing an outcome (Koopmans, 2020). This approach is common in educational research as interventions are thought to have a direct linear effect on an outcome, and analytical techniques seek to identify

these linear causes. This contradicts an idea that learning processes may be multiplicative and cannot be independently attributed to learning.

CDS methods hold significant potential for addressing these challenges. CDS draws from research in the natural sciences, including physics and thermodynamics, to understand how complex systems change and adapt. CDS methods utilize relation-intensive and time-intensive data to observe patterns of (ir)regularity across the components of the complex system (Hilpert & Manchard, 2018).

Despite the potential of CDS methods, their application to LA research is still in its infancy. This workshop aims to both educate the broader LA community about the potential of CDS, as well as help researchers who are currently applying these methods to learning data to identify common challenges in their work and transform the status quo. To enrich the LA discussion, we will invite one or two experts working in CDS but outside of learning sciences. The group discussions may revolve around key themes that may include the connections between micro-level and macro-level processes, methods that reflect within-system changes, considerations of modeling time and space, and exploration of CDS applications across various learning contexts (such as game-based learning, collaborative learning, MOOCs, intelligent tutoring systems, and hypermedia-based learning). This conversation will focus on the best practices for operationalizing models of learning that apply non-linear dynamical (NLD) analyses and common research questions that fundamentally build on the complexity science approaches for modeling various systems (e.g., individual cognition, group cognition, epistemic structure, affect dynamics, etc.). Another potential outcome is the discussion around resources, best practices, and standards in reporting NLD outcomes for easier replication.

## 3 ORGANIZATIONAL DETAILS OF THE PROPOSED EVENT

**Proposed schedule and duration**: Full day; in-person
**Type of event:** Tutorial aiming to improve general literacy around selected non-linear methods, presentations of the work-in-progress from the organizers and open discussion about scientific challenges of applying CDS to LA. The intended program is in the table below.
**Type of participation:** Any interested delegate may attend. Participants will take active part in the tutorial part of the workshop and will engage in the discussion in the second part of the workshop.

**Table 1. Workshop Organization**

| Duration | Event | Contributor |
|---|---|---|
| Part 1<br>Morning | Introduction to Complex Dynamical Systems: Basic concepts and methods. | Laura Allen, Liz Cloude, Daryn Dever |
| Part 2<br>Afternoon | Paper presentations discussed by Oleksandra Poquet and Giuseppe Leonardi, followed a keynote by Travis Wiltshire | Invited presenters |
| Part 3 | Open discussion on challenges and directions for CDS in LA | All |

**What participants should expect.** This full day workshop will consist of two distinct parts. In the first three hours of the workshop, we will present fundamental understandings of the CDSs applied to learning and engage the audience in a step-through tutorial around popular NDS methods. The focus of the tutorial will be on the aspects of quality and steps to get started, which will enhance the understanding of how these methods may be applied by the broader LA community. The second part of the workshop will consist of individual work-in-progress presentations that use NDS analysis methods applied to diverse target processes in educational research, such as in writing, self-regulated learning, emotion regulation, and social learning. Each presenter will pose major pain points and engage colleagues in discussing how these can be overcome in their work. Participants will be able to engage in these discussions and enhance their understanding of the novel techniques and the challenges associated with them.

**Expected participant numbers and planned dissemination.** Approximately 20 participants are planned. The workshop organizers are embedded in the learning analytics and related communities, and will make use of listservs (SoLAR, Learning Analytics Google group, EDM-announce, ISLS/CSCL, AERA SIG-LS, EARLI) and their networks to advertise the workshop.

## 4 WORKSHOP OBJECTIVES OR INTENDED OUTCOMES

The workshop objectives are as follows: (1) increase literacy in NLD analysis applied in learning analytics; (2) share how advanced NLD analysis methods are currently applied in various LA areas (writing, self-regulation, social learning, emotion regulation); and (3) discuss best practices and shared challenges that LA researchers collectively experience when modeling learning processes. The primary goal of the workshop is to have a productive dialogue that can align existing research efforts into a more coherent body of work accessible to the broader LA community. This is a researcher-oriented community-building workshop, hence, the underlying goal is to enable space for researchers to share and engage with one another.

The workshop website will be set up with the Github Pages, containing the workshop program and links to participants profiles. Work-in-progress presentations will be shared upon the choice of the presenter, participants will receive relevant reading and programming resources prior to the event.

## REFERENCES

Hilpert, J. C., & Marchand, G. C. (2018). Complex systems research in educational psychology: Aligning theory and method. Educational psychologist, 53(3), 185-202.

Jacobson, M. J., Kapur, M., & Reimann, P. (2016). Conceptualizing debates in learning and educational research: Toward a complex systems conceptual framework of learning. Educational psychologist, 51(2), 210-218.

Koopmans, M. (2020). Education is a complex dynamical system: Challenges for research. The Journal of Experimental Education, 88(3), 358-374.

Molenaar, P. C., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. Current directions in psychological science, 18(2), 112-117.

Poquet, O., Jovanovic, J., & Pardo, A. (2023). Student Profiles of Change in a University Course: A Complex Dynamical Systems Perspective. In LAK23: 13th International Learning Analytics and Knowledge Conference (pp. 197-207).

Richardson, M. J., Dale, R., & Marsh, K. L. (2014). Complex dynamical systems in social and personality psychology. Handbook of research methods in social and personality psychology, 253(10.1017).

# LAK24 Trace-SRL: The Workshop on Measuring and Facilitating Self-regulated Learning based on Trace data

Yizhou Fan
Peking University
fyz@pku.edu.cn

Xinyu Li
Monash University
xinyu.li@monash.edu

Mladen Raković
Monash University
mladen.rakovic@monash.edu

DraganGašević
Monash University
dragan.gasevic@monash.edu

**ABSTRACT**: This is the proposal outlines the second interactive workshop focused on Measuring and Facilitating self-regulated learning (SRL). Prior research has shown that self-regulated learning (SRL) skills are essential for successful life-long learning. Measuring SRL based on unobtrusive trace data and facilitating SRL based on real-time analysis of such trace data have been pointed out as valuable research directions. However, major challenges and significant gaps in this area are still many, such as i) the detection, measurement, and validation of SRL processes with trace data is still a much-debated issue within the SRL community; and ii) the design principles for effective interventions and the complex conditions and contexts, when these interventions facilitated learning, are still not known. Our full-day workshop is designed to advance SRL measurement and facilitation methods. Participants will gain practical experience with our Trace-SRL tools, engaging in collaborative discussions and hands-on activities to explore the application of these tools in educational settings. The workshop aims to foster a community of practice, laying the groundwork for international collaborations and furthering SRL research through shared insights, project experiences, and discussions on establishing an annual international study. Attendees will participate in roundtable discussions and co-design sessions, with the goal of catalyzing collaborative projects and laying the foundation for future joint publications.

**Keywords**: Learning analytics, Self-regulated learning, Trace data, Measurement protocols, Learning interventions, Scaffoldings and Dashboards

## 1 Background

### 1.1 Challenges

A fundamental objective within the educational domain is to cultivate Self-Regulated Learning (SRL) competencies among students. The importance of self-regulation in enhancing educational outcomes is substantiated by empirical evidence demonstrating a significant correlation between SRL processes and academic achievement (Harley, Taub, Azevedo, & Bouchet, 2017). Moreover, the capacity for self-regulation is instrumental in promoting lifelong learning (Klug et al., 2011). Nonetheless, **measuring SRL** presents a longstanding challenge within educational research. A variety of methodologies, including self-report surveys (Pintrich et al., 1991), think-aloud protocols (Bannert, 2007), and **trace-based measurement** (Siadaty et al., 2016; Fan et al., 2022), have been advanced to encapsulate SRL dynamics more accurately.

Recently, trace data analysis has emerged as a prominent method for SRL assessment (Saint et al., 2022), offering the advantage of unobtrusively capturing cognitive and metacognitive activities within genuine learning contexts. This approach aligns with the operationalization of SRL as observable learner actions (Winne, 2010) and has been employed in numerous investigations (Siadaty et al., 2016; Fan et al., 2022). Despite its growing adoption, the application of trace data in the detection, measurement, and validation of SRL processes continues to spark extensive debate among scholars in the field (Winne, 2020).

The significance of Self-Regulated Learning (SRL) in educational outcomes is well-established, yet empirical evidence consistently indicates that learners frequently struggle to effectively self-regulate across diverse settings (Azevedo et al., 2010; Winne, 2010). Despite ample opportunities for practice and refinement, SRL competencies often remain underdeveloped (Bjork et al., 2013). Consequently, there is a pressing need for supportive measures to aid learners in successfully regulating their learning processes and achieving their educational objectives. Various interventions, including **scaffolding, dashboards, and personalized feedback**, have been developed within the realm of learning analytics to bolster SRL capabilities. Nonetheless, research into the formulation of these interventions and the correlation between their design features and the facilitation of SRL, alongside learning outcomes, remains sparse (Devolder et al., 2012). Critically, the complex conditions and contexts in which these interventions effectively promote and enhance learning have yet to be elucidated (Guo, 2022).

### 1.2 Objectives

From a research perspective, this workshop aims to: i) increase awareness of how tools and data channels can be combined to measure SRL; ii) elicit new approaches for measurement and analysis of SRL; iii) understand how combining student data and artificial intelligence can be used to create actionable insights into students learning; iv) design new representations/forms of communicating SRL scaffolding, dashboards or feedback to facilitate teaching and learning. From the participant's perspective, we expect to: i) improve the knowledge and skills of participants about SRL measurement, learning processes and SRL support; ii) produce a repository of new requirements, considerations and approaches of instruments for SRL; iii) build a research community, foster partnerships, and facilitate deployment of similar tools and analytics platforms as collaborative projects; iv) explore opportunities for joint publications (e.g., a journal special issue) and future workshops resulting from the collaborations. In previous workshops, many scholars mentioned the **openness of learning platforms and tools, data sharing, and the importance of international collaborative research**. Therefore, in this year, we emphasise two objectives different from other workshops or research tracks in LAK24:

- **Provide more hands-on opportunities to experience the measuring and facilitating of SRL using our platform**. Participants will explore a learning analytic project and platform FLoRA integrated with various instrumentation tools and personalised scaffoldings, and they will be able to explore the data we provided and also the data generated by them, and then co-design possible SRL-related scaffoldings and feedback representations for learners and instructors.

- **Initiate and launch an international joint research call based on the same platform and similar tasks**. By convening researchers and educators with shared interests, our objective is to exchange insights on our platform, tasks, datasets, and project learnings, followed by deliberations on establishing an annual global collaborative research initiative. This could involve, for instance, encouraging educators to utilize a unified platform and allocate similar tasks within their courses. Such an approach enables the accumulation of data that is amenable to comparison, triangulation, and examination across various contexts, significantly enhancing the potential for international collaborative research and discourse, thereby enriching our comprehension of Self-Regulated Learning (SRL).

# 2 ORGANISATIONAL DETAILS (FULL-DAY WORKSHOP SCHEDULE)

**Table 1: Proposed Full-day Workshop Schedule (3.5 hours + 3.5 hours)**

| Timing | Descriptions | Contributors |
|---|---|---|
| | Part 1: Morning Section | |
| 20 mins | Welcome & Introduction (project and platform background) | Yizhou Fan |
| 30 mins | FLoRA 1.0 tools: annotation, timer, planner, search, essay, dictionary | Xinyu Li |
| 40 mins | Roundtable discussion 1 about improving these tools/measuring SRL | Participants |
| 30 mins | Coffee Break and Socialization | All |
| 30 mins | FLoRA 2.0 tools: scaffolding, chatgpt, chat-teacher, co-writing, checklist | Xinyu Li |
| 40 mins | Roundtable discussion 2 about improving these tools/facilititing SRL | Participants |
| 10 mins | Summarizing the morning section & Next Steps | Xinyu Li |
| | Part 2: Afternoon Section | |
| 10 mins | Welcome & Introduction (data sharing and research design) | Yizhou Fan |
| 30 mins | Data interpretation: trace parser and SRL theory models | Mladen Raković |
| 40 mins | Roundtable discussion 3 about “understanding the data we collected” | Participants |
| 30 mins | Coffee Break and Socialization | All |
| 30 mins | Ongoing development: Configure tools, data management, dashboard | Xinyu Li |
| 40 mins | Roundtable discussion 4 about designing these tools | Participants |
| 20 mins | Potential collaborations based on FLoRA platform | Yizhou Fan |
| 10 mins | Summarizing the afternoon section & Next Steps | Yizhou Fan |

The event will be an open and hands-on workshop. All attendees will have the opportunity to discuss with the organizers in the roundtable and brainstorming sessions, and will also have hands-on experiences with SRL measurement and scaffolding design activities guided by organizers. We anticipate a registration of about 10-20 participants. We will use #LAKTRACESRL when referencing this event on social media. After the workshop, **we will organize quarterly online meetings to effectively promote collaborative research, data collection and research exchanges**. And we hope to build an open, win-win and sustainable research community with the help of the LAK conferences.

We are committed to turning this workshop into an annual series of workshops and ultimately promoting in-depth exchanges and development in the field of SRL.

### 3 COMMUNICATING INFORMATION AND RESOURCES

We have a Google website and will use it to post news to our collaborators. The Google website will be the main collection point for materials, group interactions and archives for the workshop, and support ongoing dissemination and group activities. We will also disseminate information and resources about the workshop through multiple mailing lists and social media to make sure maximise the impact of the workshop.

## REFERENCES


Azevedo, R., Moos, D. C., Johnson, A. M., & Chauncey, A. D. (2010). Measuring cognitive and metacognitive regulatory processes during hypermedia learning: Issues and challenges. Educational psychologist, 45(4), 210-223.

Bannert, M. (2007). Metakognition beim lernen mit hypermedien. Waxmann Verlag.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. Annual review of psychology, 64, 417-444.

Devolder, A., van Braak, J., & Tondeur, J. (2012). Supporting self-regulated learning in computer-based learning environments: systematic review of effects of scaffolding in the domain of science education. Journal of Computer Assisted Learning, 28(6), 557-573.

Fan, Y., van der Graaf, J., Lim, L., Raković, M., Singh, S., Kilgour, J., ... & Gašević, D. (2022). Towards investigating the validity of measurement of self-regulated learning based on trace data. Metacognition and Learning, 1-39.

Guo, L. (2022). Using metacognitive prompts to enhance self-regulated learning and learning outcomes: A meta-analysis of experimental studies in computer-based learning environments. Journal of Computer Assisted Learning, 38(3), 811-832.

Harley, J. M., Taub, M., Azevedo, R., & Bouchet, F. (2017). Let's set up some subgoals: Understanding human-pedagogical agent collaborations and their implications for learning and prompt and feedback compliance. IEEE Transactions on Learning Technologies, 11(1), 54-66.

Klug, J., Ogrin, S., Keller, S., Ihringer, A., & Schmitz, B. (2011). A plea for self-regulated learning as a process: Modelling, measuring and intervening. Psychological Test and Assessment Modeling, 53(1), 51.

Pintrich, P. R., & De Groot, E. V. (1991). Motivated strategies for learning questionnaire. Journal of Educational Psychology.

Saint, J., Fan, Y., Gašević, D., & Pardo, A. (2022). Temporally-focused analytics of self-regulated learning: A systematic review of literature. Computers & Education: Artificial Intelligence, 3, 100060.

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.

Winne, P. H. (2010). Bootstrapping learner's self-regulated learning. Psychological test and assessment modeling, 52(4), 472.

Winne, P. H. (2020). Construct and consequential validity for learning analytics based on trace data. Computers in Human Behavior, 112, 106457.

# LAK24 InnovateDesign: The 2nd Workshop on Learning Design Analytics: How to Prepare AI-related Learning Design?

**Organizers 1: Blaženka Divjak, Barbi Svetec**
University of Zagreb, Faculty of Organization and Informatics, Croatia
blazenka.divjak@foi.unizg.hr, barbi.svetec@foi.unizg.hr

**Organizers 2: Dragan Gašević, Mladen Raković**
Monash University, Faculty of Information Technology, Australia
dragan.gasevic@monash.edu, mladen.rakovic@monash.edu

**Organizer 3: Bart Rienties**
The Open University, Institute of Educational Technology, The United Kingdom
bart.rienties@open.ac.uk

**ABSTRACT**: Last year's LAK conference featured a workshop dedicated to introducing an innovative concept of learning design (LD) and the utilization of a complimentary software tool for creating and analyzing LD. This year's workshop turns its focus towards the evolving challenges intertwined with AI's role in LD. Participation in last year's workshop is not a prerequisite for this year's session. The objectives for this year's workshop are twofold: To provide a platform for exchanging experiences, showcasing research findings, and deliberating on the challenges that lie at the intersection of learning analytics (LA) and LD. This encompasses the ethical and impactful integration of AI in the educational paradigm. To introduce attendees to an innovative, free LD tool (learning-design.eu) and its capabilities. Attendees are immersed in exploring advanced LD analytics using this tool. Participants are invited to collaboratively refine the LD of their courses, programs, or quality assurance endeavours while examining the LA data generated by the tool. This interactive session empowers participants to enhance their courses further, understand the role of design analytics in quality assurance, and harness the potential of AI-driven LD. This half-day, in-person workshop is a collaborative effort by three universities from Europe and Australia.

**Keywords**: learning design concept and tool, learning analytics, assessment, AI-related learning design

## 1 INTRODUCTION TO AI-RELATED LEARNING DESIGN CONCEPTS

Learning Analytics (LA) has been increasingly used to support sound learning design (LD) (Rienties et al., 2017), in particular in ensuring constructive alignment between learning outcomes (LOs), teaching and learning activities and assessment (Divjak et al., 2022, Divjak et al., 2023). LD has been defined as "the documented design and sequencing of teaching practice" (Lockyer et al., 2013, p. 1439), describing the order of learning tasks, resources and related support. LD has been presented by Conole (2013) as a methodology helping teachers and designers in more informed decision-making related to the design of learning activities, that is "pedagogically informed" and uses resources and technologies in an effective way. In essence, LD has two dimensions - a conceptual and a technological one. The concept of LD has been related to the notions of sharing and reusing: it helps make the design process "more explicit and shareable" (Conole, 2013) and aims at "reusability" in different educational contexts (Lockyer et al., 2013). It has been argued (Conole, 2013) that more explicit and shareable design enables more effective learning environments and teachers' interventions while enabling learners to understand their learning paths better. In terms of linking LD with LA, five main types of analytics have been identified in previous research: temporal, comparative, tool-specific, cohort dynamics and contingency (Bakharia et al., 2016). Today there is also a great necessity to include AI-related activities into LD in a meaningful and sound way (Dai et al., 2023; Crompton & Burke, 2023). This can, for example, include LD which supports collaboration between students and AI (Kim et al., 2022), uses AI to provide contextualized learning to students, utilizes intelligent tutoring systems, or uses AI to enhance assessment (Chaudry & Kazim, 2022). Or in other words, AI that acts as a peer and co-learning partner to students (Schoonderwoerd et al., 2022) and empowers or even mimics teachers in supporting students' acquisition of LOs (Chaudry & Kazim, 2022). Finally, AI-based LA can provide insights which can support the development and continuous improvement of LD.

Considering the recognised benefits of LD in supporting and enhancing teaching and learning in a digital age and HE teachers (Divjak et al., 2022), since 2020, a concept and a web-based tool supporting the development of sound LD, strongly supported by LA, have been developed. The Balanced Learning Design Planning (BDP) concept and tool build on the current research, theory and practice related to LD, and the existing LD concepts, primarily the OULDI approach by the Open University UK (Conole, 2013; Rienties et al., 2017), and the ABC LD approach by the University College London (Laurillard et al., 2013). The BDP concept and tool also introduce innovation in linking course LOs with the study program LOs, providing an institutional perspective. Concerning this, research has indicated that students benefit from long-term study program-level planning (Raković et al., 2022). Furthermore, the BDP tool focuses strongly on ensuring constructive alignment between LOs, types of teaching and learning activities, assessment, feedback and student workload, supporting a student-directed approach. It provides rich and deep analytics of course LD which can be used to further improve LD, in line with the intended - preferably innovative - pedagogical approaches (e.g., problem-based learning, flipped classroom, AI-related). In particular, these analytics provide detailed analyses and visualizations of assessment, minding its alignment with the prioritization, level and weights of LOs. These analytics are provided in real-time, through a dedicated dashboard, and can be used during the design process as a valuable input directing the LD process. The tool enables collaborative work and sharing of LDs, as well as the export of LDs. Finally, the tool can be used in a simple and an advanced version, enabling different levels of planning and analytics, and both versions are free to use.

At present, the BDP tool has been used in the design of more than 400 courses and MOOCs, by over 1300 users from more than 30 countries, including within four European-funded projects. Based upon the initial pilot testing (Divjak et al., 2022) and feedback from a MOOC for professional development of HE educators (Rienties et al., 2023), further functional and design modifications have been made, and at LAK 2023 we aim to share some additional functionality in terms of LA features.

## 2 LEARNING OUTCOMES, WORKSHOP STRUCTURE AND WEBSITE

Based on the capacity-building at the workshop, participants are able to (1) analyse the benefits of LA, including AI-based tools, for improvement of LD, (2) effectively use a free-to-use LD tool, including an innovative approach to LD, and (3) upgrade initial LD based on available LA. The workshop, organized in cooperation with three universities, is held face-to-face, taking half a day and consisting of the parts presented in the table below. The expected number of participants is between 15 and 30.

**Table 1. The proposed agenda of the workshop**

| Duration | Description | Responsible |
|---|---|---|
| 10 min | INTRODUCTION | Organizer 1 |
| | SHARING OF EXPERIENCES, RESEARCH AND CHALLENGES | Organizer 2 |
| 50 min | Presentations of organizers and participants' experiences | Organizer 1, 2, 3 |
| 30 min | Presentation on how the BDP can be used | Organizer 1, 2 3 |
| 30 min | BREAK | |
| | HANDS-ON COLLABORATION ON LEARNING DESIGN | |
| 70 min | Collaboration on LDs in groups working on AI-based designs | Organizer 1, 2 & 3 |
| 30 min | Presentation of LDs and discussion | Organizer 2 |
| 20 min | FUTURE STEPS AND CONCLUSIONS | Organizer 1, 2 & 3 |

The workshop is supported by a dedicated website, where all related information is shared, and which supports pre-workshop data gathering and planning, including the application of participants. To recruit participants, along with the website, social networks and media are used. After the workshop, the website and the social media are used to support ongoing dissemination. The website includes the following sections: About, Background literature and material, Workshop agenda, Submission area.

## 3 SHARING OF EXPERIENCES, RESEARCH AND CHALLENGES

The workshop starts with a few short presentations from participants and the workshop organizers, focusing on the current research, practices and experiences in the use of LD. A special focus is on how LA can support sound LD and how AI-related LD can be implemented.

Therefore, participants are invited to submit abstracts outlining short presentations (5-7 minutes) ahead of the workshop. The workshop organizers review the applications and choose interesting and diverse examples. The presentations are followed by a discussion of all participants, leading to open questions and challenges, providing introduction to the next phase of the workshop. Finally, the BDP concept and software tool are presented by the workshop organizers.

## 4 HANDS-ON COLLABORATION ON LEARNING DESIGN

Ahead of the workshop, participants are asked to consider their courses and particular LO(s) which could be redesigned at the workshop and which are suitable for AI-related teaching and learning activities. At the workshop, participants work collaboratively, grouped based on their own preferences and the similarity of courses/LOs they would like to work on.

The groups are invited to access the BDP tool, and open and design their courses and LOs focusing on AI-related activities taking into account ethical issues as well. Furthermore, they work on the detailed planning of teaching and learning activities, assessment, feedback, modes of delivery, etc. In the process, they consult the analyses provided by the tool, in order to make immediate adjustments to their LDs, aligning them with the LOs and the planned pedagogical approaches. The hands-on part of the workshop takes approximately 2 hours and each group is supported by one of the organizers. After the collaborative part, in the plenary session, groups are invited to share their LDs and mutually discuss their outputs.

# 5 FUTURE STEPS AND CONCLUSIONS

Finally, the participants are asked to take part in the evaluation of the concept and the workshop, prepared in line with the approved research protocol (ethically approved by one of the workshop organizers' universities). The conclusions of the workshop are shared with the participants after the workshop. There is a possibility to establish further collaboration to work on a project and/or a publication. All participants are able to continue using the BDP tool, as well as share it with their colleagues, free of charge.

# REFERENCES


Bakharia, A., Corrin, L., de Barba, P., Kennedy, G., Gašević, D., Mulder, R., Williams, D., Dawson, S., & Lockyer, L. (2016). A conceptual framework linking learning design with learning analytics. Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16, 329–338. https://doi.org/10.1145/2883851.2883944

Chaudhry, M. A., & Kazim, E. (2022). Artificial Intelligence in Education (AIEd): a high-level academic and industry note 2021. AI and Ethics, 2(1), 157–165. https://doi.org/10.1007/s43681-021-00074-z

Conole, G. (2013). Designing for Learning in an Open World. Springer New York. https://doi.org/10.1007/978-1-4419-8517-0

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: the state of the field. International Journal of Educational Technology in Higher Education, 20(1), 22. https://doi.org/10.1186/s41239-023-00392-8

Divjak, B., Grabar, D., Svetec, B., & Vondra, P. (2022). Balanced Learning Design Planning: Concept and Tool. Journal of Information and Organizational Sciences, 46 (2), 361-375. https://doi.org/10.31341/jios.46.2.6

Divjak, B., Svetec, B., Horvat, D., & Kadoić, N. (2023). Assessment validity and learning analytics as prerequisites for ensuring student-centred learning design. British Journal of Educational Technology, 54(1), 313–334. https://doi.org/10.1111/bjet.13290

Dai, W., Lin, J., Jin, H., Li, T., Tsai, Y. S., Gasevic, D., & Chen, G. (2023). Can large language models provide feedback to students? A case study on ChatGPT. In N-S. Chen, G. Rudolph, D. G. Sampson, M. Chang, R. Kuo, & A. Tlili (Eds.), Proceedings - 2023 IEEE International Conference on Advanced Learning Technologies, ICALT 2023 (pp. 323-325). IEEE, Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/ICALT58122.2023.00100

Kim, J., Lee, H., & Cho, Y. H. (2022). Learning design to support student-AI collaboration: perspectives of leading teachers for AI in education. Education and Information Technologies, 27(5), 6069–6104. https://doi.org/10.1007/s10639-021-10831-6

Laurillard, D., Charlton, P., Craft, B., Dimakopoulos, D., Ljubojevic, D., Magoulas, G., Masterman, E., Pujadas, R., Whitley, E. A., & Whittlestone, K. (2013). A constructionist learning environment for teachers to model learning designs. Journal of Computer Assisted Learning, 29(1), 15–30. https://doi.org/10.1111/j.1365-2729.2011.00458.x

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing Pedagogical Action: Aligning Learning Analytics with Learning Design. American Behavioral Scientist, 57(10). https://doi.org/10.1177/0002764213479367

Raković, M., Bernacki, M. L., Greene, J. A., Plumley, R. D., Hogan, K. A., Gates, K. M., & Panter, A. T. (2022). Examining the critical role of evaluation and adaptation in self-regulated learning. Contemporary Educational Psychology, 68, 102027. https://doi.org/10.1016/j.cedpsych.2021.102027

Rienties, B., Nguyen, Q., Holmes, W., & Reedy, K. (2017). A review of ten years of implementation and research in aligning learning design with learning analytics at the Open University UK. Interaction Design and Architecture(S), 33, 134–154. https://doi.org/10.55612/s-5002-033-007

Rienties, B., Divjak, B., Eichhorn, M., Iniesto, F., Saunders-Smits, G., Svetec, B., Tillmann, A., & Zizak, M. (2023). Online professional development across institutions and borders. International Journal of Educational Technology in Higher Education, 20(1), 30. https://doi.org/10.1186/s41239-023-00399-1

Schoonderwoerd, T. A. J., Zoelen, E. M. van, Bosch, K. van den, & Neerincx, M. A. (2022). Design patterns for human-AI co-learning: A wizard-of-Oz evaluation in an urban-search-and-rescue task. International Journal of Human-Computer Studies, 164, 102831. https://doi.org/10.1016/j.ijhcs.2022.102831

# Marrying Asset- and Deficit-Based Approaches: A Data Feminist Perspective in Learning Analytics

**Angela Stewart**
University of Pittsburgh
angelas@pitt.edu

**Caitlin Mills**
University of Minnesota
cmills@umn.edu

**Stephen Hutt**
University of Denver
stephen.hutt@du.edu

**ABSTRACT**: This workshop proposal explores how learning analytics can reconcile deficit- and asset-based approaches. Deficit-based models, which focus on identifying and remedying learner shortcomings, have been effective but may neglect students' existing strengths. Conversely, asset-based approaches emphasize recognizing learners' identities as assets to their learning. We advocate for a combination of both. We ground our discussion in the data feminism framework, which scrutinizes power structures in data design and interpretation. We will delve into three core data feminism principles: examine power, challenge power, and rethink binaries and hierarchies, to construct narratives affirming students' diverse identities. Through presentations, discussions, and interactive activities, we aim to develop a set of questions that allow researchers to reflect on their data and create cohesive narratives aligning asset and deficit perspectives.

**Keywords**: Learner assets, data feminism, data narratives, identity.

## 1 Workshop Background

This workshop considers how learning analytics can marry deficit- and asset-based approaches. In this context, we refer to *deficit-based* approaches as those that emphasize what learners "lack," how their performance "fails" to attain normative standards, or "gaps" between learners and their peers (Smit, 2012; Davis & Museus, 2019). Importantly, in this work we are not referring to a cultural deficit perspective which posits that students from certain groups cannot achieve due to their cultural background (Silverman, 2011). This deficit perspective ignores the larger social inequities that students from different groups experience, which is not the aim of this work, nor should it be the aim of any learning analytics work. Deficit-based models, as we define them, identify problems that might be remediated by intervention. This "find and fix" approach underpins many effective analytics because it can successfully guide students toward desired knowledge and strategies.

Though successful, we argue that this approach also incurs potential costs. Specifically, we contend that deficit-based approaches are incomplete and constrain data interpretation thus limiting the benefits that learning analytics may have. Deficit framing limits design features to reactive "filling gaps" and "fixing" students who have been (mis)categorized as "lacking," "low" knowledge or skill, "underprepared," "unmotivated," or otherwise broken. Consequently, positive and meaningful student strengths, skills, strategies, and schemas—their *assets*—are not adequately recognized or leveraged. Furthermore, this approach implicitly communicates to students (and other stakeholders) that certain assets do not matter. Meanwhile *asset-based* approaches focus on the knowledge and capabilities that learners already have, whether such strengths stem from formal educational experiences, community and cultural experiences, or family and personal life (Moll, 1992; Esteban-Guitart et al., 2014, Verdin, et al., 2021). In this workshop we argue for a combination of the two, with emphasis on recognizing students' assets, without abandoning the "find and fix" solutions that have been previously successful.

Our work is informed by the data feminism framework, which takes an intersectional approach to defining the power structures involved in designing, collecting, and interpreting data (D'ignazio & Klein, 2020). Data feminism posits that data is not neutral, and encodes elements of our identity and cultural experiences (both of those designing data collection methods, and those from whom data is collected). Particularly relevant for our context, this framework also highlights the ways data can be used to construct narratives that challenge both power structures and our understanding of students. Thus we see data feminism as being an appropriate theoretical frame for constructing cohesive asset- and deficit-based narratives that ultimately lead to actionable outcomes that better benefit students and recognise intersectional identities. In our workshop, we will focus on three of the seven core tenets of data feminism, described below:

**Examine Power.** Power refers to structural privileges or oppressions different groups experience (Collins, 2002). For example, in STEM education, women, girls, and non-binary learners often experience limited access to identity-affirming learning environments and oppressive narratives about their ability to persist in STEM fields (Scott, Sheridan, & Clarke, 2015), which is a product of patriarchy. Data feminism posits that we must examine how power operates in our world. From a learning analytics perspective, this means we examine how power operates in the lives of students, to produce different experiences outcomes for people of different identities. Additionally, this means we must examine how power structures inform our collection, analysis, and communication of data.

**Challenge Power.** Cohesive asset- and deficit-based data narratives can challenge unequal power structures by communicating understandings of data that are grounded in the lived experiences of learners. In learning analytics, this means we can use data to bring attention to the ways certain learner identities are marginalized by those in the learning space, and the social reality in which they live. For example, the work of used multimodal behavioral metrics to find that girls' in a computing camp did not verbally engage in large conversations with the instructors – a deficit narrative that focuses on what the girls' did not do compared to normative standards. However, this was married with the asset-based perspective that girls' did engage in conversations with each other in small group, student-led activities. Taken together, these analytics challenge the traditional classroom power structures that center teachers, and support student-led learning activities.

**Rethink Binaries and Hierarchies.** Binaries and hierarchies are necessary to collecting and analyzing data, as it is impossible to create data that appropriately represents the complexity of each learners' lived experience. However, as we construct asset-based narratives of learners, we must consider the ways in which binaries and hierarchies are inadequate, and confront the limits of our data. Further, we can reconsider how binaries and hierarchies uphold systems of oppression, for example by aligning with dominant views of how learners "should" show up in a learning environment. For example, categorizing "good" engagement as speaking out loud in class ignores the ways learners might be processing information non-verbally (Stewart et al. 2021). By using a data feminism framework, we will support attendees in thinking about how they can marry asset- and deficit-based perspectives. The outcome of this workshop will be a series of questions attendees can ask themself in order to better construct narratives that serve their key stakeholders.

## 2 Organization of proposed workshop

**Type of event**: Half-day hybrid workshop

**Type of participation:** The workshop will include three main components. First, we will have a workshop opening presentation, which presents the aims of the day, and discusses the data feminism framework we will use to guide our discussion. The second component will take on a mini-conference format, where a limited number of papers are presented along with discussion. For this section, we will invite research teams to submit: 1) completed projects or work-in-progress that engage in a combination of deficit- and asset-based communication or 2) opinion/commentaries that reflect on these topics and grapple with open questions about learning analytics' relationship to data feminism. Accepted paper submissions will be peer-reviewed, and accepted papers will be published in the workshop companion proceedings. The final section of the workshop will involve an interactive brainstorming activity where participants will work in small groups to come up with questions learning analytics researchers can ask themselves as they attempt to combine asset- and deficit-based narratives of learners. In this activity, participants will use an interactive whiteboard to submit questions that came up for them from the presentations earlier in the day. They will then affinity diagram (Scupin, 1997) the contributions of their group, thus drawing connections between ideas, and constructing overall themes. These themes will be converted to questions, for a wrap up discussion with the larger group. After the workshop, we will invite participants to continue the conversation on this topic by creating a special issue at an interdisciplinary journal, summarizing workshop findings and discussing how others marry asset- and deficit-based approaches.

The workshop will have open participation, with open registration for anyone interested. We expect approximately 4 - 6 paper presentations and up to 40 participants that we will recruit through website and announcements to key academic and professional communities. Some of these include: Society for Learning Analytics Research [SoLaR], Educational Data Mining Society [EDM], Society for Artificial Intelligence in Education [AIED]. Finally, we will send out targeted invitations to researchers through our personal networks, inviting both paper submissions and general attendance.

### 2.1 Workshop Format and Planned Activities

Our tentative schedule is as follows:

- 09:00am - 09:30am Workshop opening
- 09:30am - 11:00am Paper presentations + discussion
- 11:00am - 11:15am *Coffee Break*
- 11:15am - 12:15pm Interactive activity: guided small group brainstorming
- 12:30am - 12:45pm Open Discussion: Identifying guiding questions
- 12:45pm - 01:00pm Workshop closing

### 2.2 Workshop Objective and Intended Outcomes

The key outcomes of the workshop are:

1. Identify a set of probing questions that researchers can ask as they attempt to engage in cohesive asset- and deficit-based narratives that benefit students, designed to promote reflection of our role in the creation of narratives
2. Unite researchers to discuss the inevitable power structures emergent in learning data, identify ways to effectively create asset-based narratives of learners, and communicate insights to stakeholders
3. Lead a special issue at an interdisciplinary journal and summarize workshop findings, including data-driven findings as well as opinions and commentaries

## REFERENCES

Collins, Patricia Hill. Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment. New York: Routledge, 2002.

Davis, L. P., & Museus, S. D. (2019). What is deficit thinking? An analysis of conceptualizations of deficit thinking and implications for scholarly research. NCID Currents, 1(1).

D'ignazio, C., & Klein, L. F. (2020). Data feminism. MIT press.

Scott, K. A., Sheridan, K. M., & Clark, K. (2015). Culturally responsive computing: A theory revisited. *Learning, Media and Technology*, *40*(4), 412-436.

Esteban-Guitart, M., & Moll, L. C. (2014). Funds of identity: A new concept based on the funds of knowledge approach. Culture & Psychology, 20(1), 31-48.

Moll, L. C. (Ed.). (1992). Vygotsky and education: Instructional implications and applications of sociohistorical psychology. Cambridge University Press.

Scupin, R. (1997). The KJ method: A technique for analyzing data derived from Japanese ethnology. *Human organization*, *56*(2), 233-237.

Silverman, S.K. (2011). Cultural Deficit Perspective. In: Goldstein, S., Naglieri, J.A. (eds) Encyclopedia of Child Behavior and Development. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-79061-9_750

Smit, R. (2012). Towards a clearer understanding of student disadvantage in higher education: Problematising deficit thinking. Higher Education Research & Development, 31(3), 369-380.

Stewart, A. E. B., Solyst, J., Buddemeyer, A., Hatley, L., Henderson-Singer, S., Scott, K., Walker, E., & Ogan, A. (2021). Explaining Engagement: Learner Behaviors in a Virtual Coding Camp. In I. Roll, D. McNamara, S. Sosnovsky, R. Luckin, & V. Dimitrova (Eds.), Artificial Intelligence in Education (pp. 338–343). Springer International Publishing. https://doi.org/10.1007/978-3-030-78270-2_60

Verdín, D., Smith, J. M., & Lucena, J. (2021). Funds of knowledge as pre-college experiences that promote minoritized students' interest, self-efficacy beliefs, and choice of majoring in engineering. Journal of Pre-College Engineering Education Research (J-PEER), 11(1).

# CELLA@LAK24 Workshop: Supporting Students' Self-Regulated Learning Through Human-AI Collaboration

**Joni Lämsä[1], Susanne de Mooij[2], Lyn Lim[3], Daryn Dever[4], Megan Wiedbusch[4], Roger Azevedo[4], Maria Bannert[3], Dragan Gasevic[5], Inge Molenaar[2], & Sanna Järvelä[1]**
[1]University of Oulu, Finland; [2]Radboud University, Netherlands; [3]Technical University of Munich, Germany; [4]University of Central Florida, USA; [5]Monash University, Australia

**ABSTRACT**: Our international, multidisciplinary research Centre for Learning and Living with AI (CELLA) organises this interactive workshop on how to use multiple data streams to measure and support students' self-regulated learning (SRL) through human-AI collaboration. Prior research has shown that supporting SRL through learning analytics (LA) fosters life-long learning skills. However, there are still major challenges for the LA community conducting research in this area: i) identifying useful data streams to measure different SRL processes in an unobtrusive, valid, and reliable manner; and ii) supporting SRL with LA backed interventions. Therefore, this full-day workshop facilitates a program of research integrating different types of SRL trace data into LA-based supports by i) presenting empirical and theoretical studies; ii) initiating multidisciplinary dialogue (e.g., computer science, learning sciences) to foster cross-team collaborations and promote transdisciplinary perspectives on human-AI collaborations for SRL; and iii) providing workshop participants with hands-on opportunities to collect multi-trace data and investigate different personalized support types (e.g. dashboard, scaffolding, NLP generated prompts) based on human-AI collaboration. Expected outcomes are forming a community of research and practice within the field LA; identifying potential areas for collaborative projects; and promoting future collaborations for joint publications and grant submissions.

**Keywords**: Learning analytics, Self-regulated learning, Trace data, Measurement protocols, Learning interventions, Adaptive scaffolds, Learning dashboards, Human-AI collaboration

## 1 BACKGROUND

### 1.1 Challenges

To navigate through the faster pace of life and work in the age of artificial intelligence (AI) students are required to develop self-regulated learning (SRL) skills to monitor and control their cognition, affect, metacognition, and motivation during learning (Järvelä et al., 2023). Despite the recognized benefits of SRL and the numerous opportunities for students to enhance these skills, students' SRL skills remain underdeveloped (Azevedo & Wiedbusch, 2023). While research has investigated the support of students' SRL through various advanced learning technologies (e.g., intelligent tutoring systems, simulations, hypermedia, serious games, virtual reality), it often overlooks interventions based on LA insights to enhance learning (Ferguson et al., 2023). In agreement with others (e.g., Gašević et al., 2023), we emphasize the need for LA research to focus on student learning, closing the loop from detection to intervention within a single research program (e.g., Ferguson et al., 2023).

Previous studies on supporting SRL based on rule-based AI and automatic labelling of SRL processes based on digital multi-trace data have shown effects on the learning process but more work still needs to be done to optimize how students use the scaffolds (Lim et al., 2023). In addition to scaffolding students *during* learning, support for SRL can be provided *after* learning with the help of dashboards (Matcha et al., 2019). Although SRL dashboards have potential to provide students insight into their own learning, more work has to be done to make them theoretically grounded and actionable (Matcha et al., 2019). The third approach to support SRL is to provide training *before* learning. For example, targeted SRL training on metacognitive activities can improve learning performance (Bannert et al., 2008). Recent AI and large language models (LLMs) advancements offer new avenues for SRL support through human-AI collaboration. In the context of supporting SRL, we see human-AI collaboration as a form of shared regulation (Järvelä et al., 2023). Understanding human-AI regulation as a reciprocal interaction, in which AI considers human conditions to tailor support, fosters the learner's ability to regulate. This interactive process benefits both humans and AI, facilitating the transition from adaptive learning to empowering students who can self-regulate their learning with the help of AI.

Regardless of support type, the reliable and valid measurement of SRL forms the basis for supporting SRL with AI in a timely and adequate manner. Due to the covert and intertwined nature of cognitive, affective, motivational, and metacognitive processes, the measurement of SRL has been a major challenge, and researchers have increasingly applied multiple data streams to capture these processes through integrated approaches using machine learning techniques (Molenaar et al., 2022). However, the challenge remains that even though these multiple data streams are feasible to collect in laboratory settings, their transfer to authentic classroom settings is problematic regarding accessibility and quality control. Thus, it is important to identify the strengths and limitations of collecting and using multimodal data for identifying how students learn with AI-based technologies.

## 1.2 Objectives

From a research perspective, this workshop aims to: i) increase awareness of how different data streams can be combined to measure SRL in an unobtrusive, valid, and reliable manner across contexts (lab and classroom settings); ii) elicit new approaches for supporting SRL before, during, and after learning while fostering learners' agency; iii) understand the different forms of human-AI collaboration that can be initiated when supporting SRL; and iv) demonstrate how combining data on learning processes and AI can be used to create actionable insights into and facilitate students' learning with different representations (e.g., dashboards or ChatGPT generated prompts).

From the participants' perspective, we expect to: i) improve the knowledge and skills of participants about challenges in SRL measurement; ii) provide a repertoire of approaches to support SRL with an emphasis on support through human-AI collaboration; iii) build a research community, foster partnerships, and facilitate deployment of tools and analytics platforms as collaborative projects; and iv) explore opportunities for joint publications, grants, and future workshops resulting from the collaborations. The outcomes of the workshop will be housed on a Google Site. More specifically, we have two objectives for this workshop:

- Initiate a project-to-project level dialogue to foster cross-team collaboration. By bringing together projects that measure learning and learning processes with different analytical approaches and deliver support in the forms of human-AI collaboration, we aim to promote research and practice that look at learning from a more comprehensive perspective than single studies.
- Provide hands-on opportunities to experience different analytical approaches to measuring SRL and working with different SRL support tools. These measures and support SRL tools will be conducted with an existing platform with various instrumentation tools and support options, including possibilities for human-AI collaboration. Participants will be able to explore their own multi-channel data and SRL-related support and feedback representations.

# 2 ORGANISATIONAL DETAILS (FULL-DAY WORKSHOP SCHEDULE)

| Timing | Description | Host(s) |
|---|---|---|
| | Part 1: Morning Section: Measuring SRL | |
| 9:00-9:10 | Welcome & Introduction | Susanne de Mooij, Joni Lämsä |
| 9:10-9:40 | Towards human-AI collaboration in measuring and supporting self-regulated learning | Sanna Järvelä |
| 9:40-10:40 | 1. A Systematic Review of Measurement of Self-Regulated Learning Through Integration of Multimodal Data and AI (de Mooij et al.)<br>2. The Generative Multimodal Analysis (GMA) Methodology for Studying Socially Shared Regulation in Collaborative Learning (Whitehead et al.)<br>3. Measuring Self-regulated Learning with Learning Trace Data – Mapping Theoretical Construct to Traces (Li et al.) | Chair: Joni Lämsä |
| | Coffee break & socialisation | |
| 11:15-12:00 | Hands-on task to create own SRL data and presentation of collected multimodal data | Xinyu Li |
| 12:00-12:15 | Summarising the morning section and next steps | Dragan Gasevic |
| | Part 2: Afternoon section: Supporting SRL through human-AI collaboration | |
| 13:45-14.00 | Introduction (afternoon session) | Susanne de Mooij, Joni Lämsä |
| 14:00-15:00 | 1. Enhancing Self-Regulated Learning through Theory-Based Prompting and Large Language Models: Insights from Medical Education (Stadler et al.)<br>2. Improving Self-Regulated Learning through Theoretically Driven Rule-Based AI Personalized Scaffolds: Implications for Optimizing Scaffolds (Lim et al.)<br>3. The Role of Learning Analytics in Human Digital Twins: From Theory and Design to Collaborative Learning Applications (Wiedbusch et al.) | Chair: Maria Bannert |
| | Coffee break & socialization | |
| 15:30-16:30 | Hands-on workshop with personalized support type tools<br>1. Training SRL (before learning)<br>2. Scaffolding and prompts (during learning)<br>3. Dashboard (after learning) | <br>1. Lyn Lim<br>2. Daryn Dever<br>3. Susanne de Mooij |
| 16:30-16:50 | Open discussion on implications and new directions of support types before, during and after learning | Chair: Roger Azevedo |
| 16:50-17:00 | Summarizing the afternoon section and next steps | Sanna Järvelä |

The event will be a hands-on workshop. The organization of the workshop will revolve around cutting-edge research projects related to measuring and supporting SRL through human-AI collaboration, so we will collect research abstracts as the basis for the workshop. Abstract submissions of 500 words for these projects will be handled via the workshop's website, and each submission should contain both the measuring or supporting of SRL as well as a component of human-AI collaboration; both work in progress and completed studies are considered as valuable contributions to this workshop, given the novelty of the human-AI collaboration aspect in the field of supporting SRL. The main purpose of this arrangement is to make the two sections in the morning (on measuring SRL in the lab and field) and afternoon (support through human-AI collaboration before, during and after learning) to echo each other and provide workshop participants with a global understanding and in-depth discussion about SRL. The submission timeline will follow the timeline suggested by the conference organizers. All attendees will have the opportunity to discuss with the presenters and will also have hands-on experiences with SRL measurement and support tools guided by organizers.

## 3 COMMUNICATING INFORMATION AND RESOURCES

We will have a Google website and will use it to post the call-for-papers. At the same time, we will send invitations to specific relevant research teams. The Google website will be the main collection point for materials, group interactions and archives for the workshop. We will also disseminate information and resources about the workshop through multiple mailing lists and social media to make sure to maximise the impact of workshop.

## REFERENCES

Azevedo, R., & Wiedbusch, M. (2023). Theories of metacognition and pedagogy applied to AIED systems. In B. du Boulay, A. Mitrovic, & K. Yacef (Eds.), Handbook of Artificial Intelligence in Education (pp. 45–67). Edward Elgar Publishing.

Ferguson, R., Khosravi, H., Kovanović, V., Viberg, O., Aggarwal, A., Brinkhuis, M., ... & Yan, V. X. (2023). Aligning the Goals of Learning Analytics with its Research Scholarship: An Open Peer Commentary Approach. Journal of Learning Analytics, 10(2), 14-50.

Gašević, D., Siemens, G., & Sadiq, S. (2023). Empowering learners for the age of artificial intelligence. Computers and Education: Artificial Intelligence, 4, 100130.

Järvelä, S., Nguyen, A., & Hadwin, A.F. (2023). Human and AI collaboration for socially shared regulation in learning. British Journal of Educational Technology, 54(5), 1057-1076.

Lim, L., Bannert, M., Graaf, J. van der, Singh, S., Fan, Y., Surendrannair, S., Rakovic, M., Molenaar, I., Moore, J., & Gašević, D. (2023). Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning. Computers in Human Behavior, 139, 107547.

Matcha, W., Gašević, 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.

Molenaar, I., de Mooij, S., Azevedo, R., Bannert, M., Järvelä, S., & Gašević, D. (2022). Measuring self-regulated learning and the role of AI: Five years of research using multimodal multichannel data. Computers in Human Behavior, 139, 107540.

# Advancing Actionability in Learning Analytics by Uniting Diverse Stakeholder Perspectives

**Yeonji Jung[1], Alyssa Wise[2], Yannis Dimitriadis[3], Ishari Amarasinghe[4]**
University of Memphis, yeonji.jung@memphis.edu
Vanderbilt University, alyssa.wise@vanderbilt.edu
Universidad de Valladolid, yannis@tel.uva.es
Radboud University, ishari.amarasinghe@donders.ru.nl

**Abstract:** The pursuit of actionability in learning analytics has long been a central aim, yet the knowledge base related to improving it has remained relatively sparse and disconnected. This workshop aims to initiate unifying discussions on how "actionability" can be conceptualized for the learning analytics community. During the workshop, we will define and refine actionability from various stakeholder perspectives: technical (for tech developers), design (for designers), self-regulated learning (for learners), and classroom orchestration (for teachers); and then explore how these perspectives can be used to inform the development of analytics tools, learning designs, and impact measurement. Through diverse discussions and consolidation efforts, this workshop seeks to develop a comprehensive framework with tangible implications and foster a network of interested researchers and practitioners in actionable learning analytics.

**Keywords**: Actionability, learning analytics, human-centered design, orchestration, self-regulated learning

## 1 WORKSHOP OBJECTIVES

### 1.1 Background and Relevance

The intent to offer insights into learning that are "actionable" has been a core tenet of learning analytics from the field's inception (Siemens, 2013). That this goal of making an impact on practice remains more aspirational than realized has been noted by multiple papers, both those examining prior impact (Ferguson et al., 2016) and those considering how the situation might be rectified (Dimitriadis et al., 2021). Increased attention to developing learning analytics that are not only technically rigorous but able to be effectively used by teachers, students, and other educational stakeholders to improve learning can be seen as part of a larger move toward human-centered learning analytics that takes people's situations, needs, and goals as the starting point (Buckingham Shum, Ferguson, & Martinez-Maldonado, 2019). For example, recent work has started to bring actionability to the forefront, anchoring it as part of fundamental inquiry for impactful learning analytics research (Dimitriadis et al., 2021; Jørnø & Gynther, 2018; Jung & Wise, 2024).

While the importance of the actionability of learning analytics may be acknowledged widely, the knowledge base related to improving it has remained relatively sparse and disconnected. This may be due both to the fact that many researchers do not have the opportunity to concretely address actionability in authentic learning situations and that those who do have approached the challenge in quite different ways. For example, harkening back to the "actionable insights" language of business analytics, some have focused on the effective presentation of useful information that can be acted upon to improve learning (Susnjak, Ramaswami, & Mathrani, 2022). From this perspective, there is a growing interest in tool design that ensures that metrics are not only informative but also motivational to prompt particular actions (Dimitriadis et al., 2021). Work taking this view often emphasizes human-centered design approaches, involving stakeholders in design to identify actual needs and preferences that can inform the design decisions (Wiley, Dimitriadis, & Linn, 2023). A related approach to actionability involves embedding actions or links to action in analytics tools; for example, pre-written messages from instructors can be programmed to send automatically to students with certain levels of activity in an online system as tracked by the analytics (Pardo, Jovanovic, Dawson, Gašević & Mirriahi, 2019). In contrast to the technical perspectives described above, other work has emphasized the social aspects of actionability, considering end-user routines and integration of analytics into their practices. From these perspectives, actionability depends not only on the types of information provided and visualization cues but on broader social systems of activity, taking into consideration such factors as teacher orchestration and student self-regulated learning (Amarasinghe et al., 2022; Klein et al., 2019). In this view, actionability is not simply a property of the analytics but also the larger learning system into which they are embedded. This allows for a broader perspective on the impact of analytics in teaching and learning activities, including both direct behaviors and decisions based on analytics, as well as more holistic or implicit ways that analytic information can feed into the ways these systems operate (Wise & Jung, 2019).

### 1.2 Objectives and Outputs

This workshop will initiate unifying discussions on how “actionability” can be conceptualized for the learning analytics community; first articulating and refining the concept of actionability in learning analytics from different stakeholders' perspectives; and then exploring how these perspectives can be used to inform (a) the creation of learning analytics tools, (b) learning designs that incorporate such tools, and (c) measurement of learning analytics impact. During this process, we will strive to initiate divergent discussions, later incorporating various aspects of actionability into a consolidated framework with concrete implications. This workshop is expected to produce (1) a clear articulation of the consolidated conceptualization of actionability with implications for learning analytics research and design (2) a written artifact that will be published for dissemination and feedback from the community (e.g., a mailing list) and (3) a network of researchers and practitioners.

## 2 ORGANIZATIONAL DETAILS

### 2.1 Duration and Format of Event

The workshop is planned to be a half-day, face-to-face event. The workshop will take the format of interactive workshop, where a maximum of 20 participants with a shared interest in actionable learning analytics are expected to engage in presentations and small-group discussions.

### 2.2 Call for Papers and Pre-Workshop Tasks

A call for short (2-4pp) papers related to any of the perspectives on actionability described above (information design, tool functionality, teacher orchestration, student self-regulation) or additional ones identified by participants (open call) will be released and circulated via relevant listservs (e.g., Learning Analytics google group, International Society of the Learning Sciences listserv) and personal networks. Submitted papers will be peer-reviewed by the organizers and those submitting papers. A maximum of six papers will be selected based on their relevance and contribution to different perspectives on actionability. Participants will also be able to attend as regular (not presenting) contributors; they will be asked to complete a survey before the workshop, which will ask about their experience and interest in the different perspectives on actionability.

## 2.3 Planned Activities

**Part 1. Introduction to the workshop**

- Short introductory presentation by the organizers about the workshop theme and goals, and overall (implicit/unclear) definitions of actionability with respect to related areas of research in the field such as human-centered learning analytics and teacher orchestration.
- **Introductory presentation** by the organizers, sharing the concept of actionability in learning analytics from different stakeholders' perspectives based on the literature synthesis: (1) technical aspect (for technology designers and developers), (2) information presentation (for designers), (3) self-regulated learning (for learners), (4) classroom orchestration (for teachers). This set of perspectives will be used as a starting point for discussing the concept of actionability.

**Part 2. Different perspectives on actionability for different stakeholders**

- **Brief talks** presented by the accepted paper presenters about their research experiences related to each of the perspectives on actionability.
- Participants will follow a World Café script, engaging in multiple rounds of small-group activities, each with a specific question. Four groups, aligned with the four main perspectives, will be formed (if additional perspectives are identified, additional groups can be added and a faster rotation speed introduced). Each 15-minute round allows participants to switch tables and discuss a different perspective. Organizers, acting as table hosts, will facilitate each round, welcoming new groups to the table and summarizing previous discussions on the assigned perspective. To facilitate group work, the Miro software (https://miro.com) will be used for brainstorming and synthesis of shared insights.
- **Group work for specifying conceptualization and producing implications:** Each round will start by asking participants to collaborate on a shared whiteboard to refine the concept of actionability from each of the perspectives. They will also identify key considerations, issues, and contributing factors for each perspective, incorporating them into specifications. Then, participants will brainstorm how each of the perspectives can be used to inform: (a) the creation of learning analytics tools, (b) learning designs that incorporate such tools, and (c) measurement of learning analytics impacts. During this activity, participants will be provided with 1-2 concrete examples of learning scenarios that they can relate to.
- **Sharing conceptualization and implications across groups:** Once the group work is done, the organizers will ask participants to look at the synthesized work from the prior steps across the perspectives. All attendees may engage in a discussion to integrate and/or identify points of synergy and tension among the multiple perspectives.

**Part 4. Discussions on Next Step**

- The workshop will conclude with a return to the first part. Participants will be invited to collectively brainstorm potential research directions and/or practical steps based on their workshop participation. They will be asked to fill out their contact information, research agenda, and willingness for potential collaboration in a co-working document that others can view and edit together.

### 2.3 Expected Outcome and Dissemination

The outcomes of the workshop will be published as an open-access report for the wider community, while disseminated to all workshop participants using the mailing lists. This report will synthesize key outcomes including conceptualizations, considerations, and implications of actionability as well as the accepted papers.

### 2.4 Communication and Dissemination

An online Google website will be created and used for posting a call for papers and informing potential participants of relevant information. The call for papers will be published via the website, and direct invitation contact to specific research teams whose works have been addressing actionability in learning analytics. The website will serve as a central repository for materials, group interactions, workshop archives, and ongoing dissemination and networking.

## REFERENCES

Amarasinghe, I., Michos, K., Crespi, F., & Hernández-Leo, D. (2022). Learning analytics support to teachers' design and orchestrating tasks. Journal of Computer Assisted Learning. DOI: 10.1111/jcal.12711

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

Dimitriadis, Y., Martínez-Maldonado, R., & Wiley, K. (2021). Human-centered design principles for actionable learning analytics. Research on E-learning and ICT in education: Technological, pedagogical and instructional perspectives, 277-296.

Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., Rienties, B., Ullmann, T., & Vuorikari, R. (2016). Research evidence on the use of learning analytics: Implications for education policy. R. Vuorikari, J. Castaño Muñoz (Eds.). Joint Research Centre Science for Policy Report; EUR 28294 EN; doi:10.2791/955210.

Jørnø, R. L., & Gynther, K. (2018). What constitutes an 'actionable insight'in learning analytics?. Journal of Learning Analytics, 5(3), 198-221.

Jung, Y. & Wise, A. F. (2024). Probing actionability in learning analytics: The role of routines, timing, and pathways. In Proceedings of LAK 2024. 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.

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.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380-1400.

Susnjak, T., Ramaswami, G. S., & Mathrani, A. (2022). Learning analytics dashboard: a tool for providing actionable insights to learners. *International Journal of Educational Technology in Higher Education, 19*(1), 12.

Wiley, K., Dimitriadis, Y., & Linn, M. (2023). A human-centred learning analytics approach for developing contextually scalable K-12 teacher dashboards. *British Journal of Educational Technology*. https://doi.org/10.1111/bjet.13383

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.

# "Social Network Analysis for Newbies: Theory, Applications, and Analysis"

**Shaun Kellogg, Jeanne McClure, Daria Smyslova, Shiyan Jiang, Doreen Mushi,**
North Carolina State University
sbkellog@ncsu.edu, jmmcclu3@ncsu.edu, dysmyslo@ncsu.edu,
sjiang24@ncsu.edu, drmushi@ncsu.edu

**Susan Hibbard**
BluePrintPrep
susan.hibbard@blueprintprep.com

**ABSTRACT**: The resurgence of interest in social network analysis (SNA) in educational contexts, spurred on by the proliferation of social networking sites and the integration of digital resources in education, is the focal point of this one-day course. Aimed at education researchers unfamiliar with SNA, the course offers an immersion into social network theory, showcases diverse applications of network analysis in educational settings, and affords hands-on experience with analyzing actual data sets. By weaving theoretical instruction with applied experiences, the course seeks to foster a deep-seated understanding of SNA's dual role as a theoretical lens and a method of analysis, enabling scholars to harness its potential in understanding and enhancing learning environments and outcomes. This endeavor champions the belief that SNA can be a powerful tool in the continuous effort to improve student learning and the atmospheres in which this learning takes place.

**Keywords**: social network analysis, social capital, educational patterns, learning analytics

## 1 BACKGROUND:

Social Network Analysis (SNA) has a rich history dating back to the early 20th century, where it was utilized to explore school friendships and other educational phenomena (Moreno & Jennings, 1938; Wellman, 1926; Almack, 1922). In recent decades, there has been a notable surge in scholarly attention towards SNA, mirroring a broader academic shift towards more relational and systemic approaches in understanding complex phenomena, and steering away from individual-centric perspectives (Borgatti & Foster, 2003; Saqr, Poquet, Lopez-Pernas, 2022).

Echoing this trend, advancements in computing power and statistical methodologies have propelled a significant uptake in network research in education, powered further by the increased visibility and understanding of networks brought forth by social media platforms (Mcfarland, Diehl, & Rawlings, 2011). Scholars are increasingly leveraging Social Network Analysis (SNA) to provide rich and multifaceted perspectives on educational environments. They explore various aspects such as social capital, peer influences, and the diffusion of innovations, showcasing the extensive potential applications of SNA for understanding and improving educational settings (Mcfarland, Diehl, & Rawlings, 2011; Carolan, 2013).

In addition to measuring communities in discussion forums, the field of education has witnessed a rising application of SNA methods. These methods are employed to understand actors and actor groups engaged in debates about educational inclusivity (Schuster et al., 2021), investigate patterns of collaboration and advice-seeking among schools (Sinnema et al., 2020), and explore how social presence in an ongoing course evolves with hardcoded discussions and log data (Norz et al., 2023).

The Learning Analytics in STEM Education Research (LASER) Institute was developed with the primary goal of increasing the number and the capacity of scholars capable of leveraging new data sources and computational methods (e.g., network analysis, text mining and machine learning) to support their research. This half-day SNA workshop serves as an extension of the LASER Institute and is designed for a diverse range of education researchers, including early-career researchers, PhD students, faculty members, and practitioners, who are keen on exploring innovative methodologies to enhance their research.

Using LASER curriculum materials, the half-day course offers an immersion into social network theory, showcases diverse applications of network analysis in educational settings, and affords hands-on experience with analyzing actual data sets. SNA is pertinent to foster a deeper understanding of its multifaceted applications, grounded in rich historical context, and to equip attendees with the nuanced perspective required to navigate the complex landscapes of contemporary educational environments. It serves as a conduit to delve deeper into the transformative potential SNA holds in scrutinizing and enriching learning analytics from a systemic and contextual standpoint, tracing a trajectory that has evolved significantly from its inception.

By combining an introduction to network theory with applied experience conducting network analyses, the workshop aims to improve participating scholars' understanding of SNA's dual role as a theoretical lens and a method of analysis. This balanced approach enables scholars to harness the potential of SNA in understanding and enhancing learning environments and outcomes. The workshop, designed to address the current needs and interests of educational researchers, serves as a comprehensive introduction to the transformative potential of SNA in learning analytics. Attendees will leave with a comprehensive understanding of the theory and application of SNA, practical skills in analyzing educational networks, and access to resources and materials to support their ongoing learning and research.

# 2 ORGANIZATION:

## 2.1 Type of Event

This session will be an interactive workshop.

## 2.2 Duration

The workshop will follow a half day format.

## 2.3 Workshop activities

The workshop will include presentation, a guided activity, small- and large-group discussions and a hands-on activity.

# 3 OBJECTIVES AND OUTCOMES

Broadly, this workshop offers those in the Learning Analytics community an exposure to an introduction of SNA for Learning Analytics. The objective of this course is to facilitate scholars in an introduction to the robustness of SNA not just as an alternative but also a supplementary method to the conventional research techniques. The detailed learning goals for participants are outlined as follows:

- **Theory Comprehension:** Acquire knowledge on the theoretical underpinning of social network analysis, and understand its application in solving critical problems and addressing pertinent questions in the educational sector.
- **Identifying Data and Metrics**: Learn to pinpoint potential data sources for network analysis, and familiarize oneself with related metrics such as centrality and degree.
- **Software Mastery:** Become adept at utilizing current software and tools like R and Python, enhancing skills in the execution of workflows for data preparation, analysis, and dissemination.
- **Analytical Understanding:** Grasp the analytical procedures and techniques such as sociograms and clustering in network analysis, essential for comprehending and augmenting learning as well as the environments conducive to learning.
- **Effective Communication:** Develop an understanding of the fundamental concepts and terms in SNA, empowering individuals to convey basic SNA methods, analytical outcomes, and discoveries to a broad spectrum of stakeholders in education.

Although having a background in R, RStudio, and GitHub can aid in navigating complex activities, it is not mandatory.

# 4 COMMUNICATION PLAN

## 4.1 Recruitment

The organizers will recruit through individual invitations, social media platforms, networks, the Learning Analytics Google Group, and the conference website.

## 4.2 Information Sharing

The organizers will communicate via email prior to and following the event. The workshop organizers will create a welcome packet to distribute to participants prior to the workshop. This packet will contain essential materials, including pre-workshop recommendations, and information about the technology tools that will be utilized both during and after the workshop.

## 4.3 Recruitment

The organizers plan to make use of a website, Posit Cloud and Github repository.

## REFERENCES

Almack, J. C. (1922). The influence of intelligence on the selection of associates. School and Society, 16(410), 529-530.

Borgatti, S. P., & Foster, P. C. (2003). The network paradigm in organizational research: A review and typology. Journal of management, 29(6), 991-1013.

Carolan, B. V. (2013). Social network analysis and education: Theory, methods & applications. Sage Publications

McFarland, D. A., Diehl, D., & Rawlings, C. (2011). Methodological transactionalism and the sociology of education. In Frontiers in sociology of education (pp. 87-109). Dordrecht: Springer Netherlands.

Moreno, J. L., & Jennings, H. H. (1938). Statistics of social configurations. Sociometry, 342-374

Norz, L. M., Dornauer, V., Hackl, W. O., & Ammenwerth, E. (2023). Measuring social presence in online-based learning: An exploratory path analysis using log data and social network analysis. The Internet and Higher Education, 56, 100894.

Saqr, M., Poquet, O., & López-Pernas, S. (2022). Networks in education: A travelogue through five decades. IEEE Access, 10, 32361-32380.

Schuster, J., Jörgens, H., & Kolleck, N. (2021). The rise of global policy networks in education: analyzing Twitter debates on inclusive education using social network analysis. Journal of Education Policy, 36(2), 211-231.

Sinnema, C., Daly, A. J., Liou, Y. H., & Rodway, J. (2020). Exploring the communities of learning policy in New Zealand using social network analysis: A case study of leadership, expertise, and networks. International Journal of Educational Research, 99, 101492.

Wellman, B. (1926). The school child’s choice of companions. The Journal of Educational Research, 14(2), 126-132.

## ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. DRL-2025090 and DRL-2321128 for the Learning Analytics in STEM Education Research (LASER) Institute. The authors would also like to thank the following members of the LASER team for their invaluable contributions: Co-PI Rob Moore and Senior Personnel Joshua Rosenberg, Jennifer Houchins and the support team members. The authors are also grateful to the over 120 scholars who have participated in this professional learning program and have helped to improve LASER through their active engagement and ongoing feedback.

# Learning to Teach with (and Learn from!) the LASER Curriculum

**Shaun Kellogg, Jeanne McClure, Daria Smyslova, Shiyan Jiang, Doreen Mushi,**
North Carolina State University
sbkellog@ncsu.edu, jmmcclu3@ncsu.edu, dysmyslo@ncsu.edu, sjiang24@ncsu.edu, drmushi@ncsu.edu

**Susan Hibbard**
BluePrintPrep
susan.hibbard@blueprintprep.com

**Ryan Baker**
University of Pennsylvania
ryanshaunbaker@gmail.com

**ABSTRACT**: The purpose of this interactive workshop is to provide a hands-on introduction to curriculum developed as part of the Anonymized Institute, a professional development program for early and mid-career researchers funded by the National Science Foundation (ECR: BCSER). The intended audience for this workshop includes early-career and experienced scholars seeking who currently teach, or have a desire to teach, learning analytics methodologies. The primary aim of this workshop is to support participants interested in incorporating Anonymized curriculum materials into webinars, workshops, courses or programs at their home institution. Participants in this workshop will learn about the design and structure of the 25+ learning modules covering a range of topics and techniques like machine learning, network analysis, and text mining. Participants will also gain hands-on experience with instructional activities such as conceptual overviews, interactive code-alongs, tutorials, case studies using Python and R, essential readings and discussion activities, and badging and microcredential opportunities. Finally, participants will learn pedagogical tips and information on the computing infrastructure, technology stack, and logistics required for leveraging these materials for their own undergraduate, graduate or professional learning programs.

**Keywords**: graduate education, professional learning, professional development, STEM education, machine learning, network analysis, text mining, relationship mining, knowledge tracing, microcredentials

## 1 OBJECTIVES

### 1.1 Introduction

In response to the rapid digitization of teaching and learning, Learning Analytics (LA) has emerged over the past 15 years as an interdisciplinary practice for understanding and optimizing the measurement, analysis, and reporting of student data to better understand and improve the contexts in which they learn (Means & Anderson, 2013; Alexander et al., 2019; Krumm et al., 2018; Siemens, 2014). Despite the rich opportunities for STEM education research afforded by LA, there are still relatively few academic programs in Learning Analytics or related fields (e.g. educational data science) and "most educational research programs do not require data literacy beyond a graduate statistics course" (Dede, 2016, p. 110). While a growing number of general data science courses, bootcamps, and other offerings have helped to fill this capacity-building gap, these offerings are disconnected from the context of education and applications specific to STEM teaching and learning.

To help address this need for researchers trained in LA and related methods, the Learning Analytics in STEM Education Research (LASER) Institute was developed with the primary goal of increasing the number and the capacity of scholars capable of leveraging new data sources and computational methods (e.g., network analysis, text mining and machine learning) to support their research. With a new round of funding from the National Science Foundation (ECR: BCSER), North Carolina State University and the University of Pennsylvania are refining, expanding, and repackaging instructional resources developed for the LASER Institute into “turnkey” curriculum materials that can be used and adapted by faculty in higher education to prepare the next generation of STEM scholars.

### 1.2 Workshop Goals

The half-day workshop serves as an extension of the LASER Institute, with the primary goal of equipping participants with high-quality curriculum materials that can be used to train faculty and students at their home institutions. To achieve this goal, this workshop is organized into three parts. Part 1 introduces participants to the curriculum materials developed by the LASER team and provides hands-on activities to help participants understand curriculum content and instructional design. Part 2 focuses on the technology infrastructure required for teaching with these materials including set up for Posit Cloud, R Studio, and Jupyter Notebooks. In addition, facilitators will help participants select appropriate modules for their instructional goals and develop a tentative plan to use curriculum materials for webinars, workshops, and courses at their university or research institutions.

Finally, we recognize that there is always room for improving these curriculum materials, particularly after use in a wider range of instructional settings. Therefore, a secondary aim of this workshop is to gather feedback from participating scholars both during and after the workshop to further refine the curriculum materials. Specifically, we will incorporate opportunities throughout the workshop to solicit feedback from participants on how curriculum materials might be improved or adapted to better fit their local context and meet the needs of learners at their home institutions.

## 2 ORGANIZATIONAL DETAILS

### 2.1 Workshop Activities

The workshop will kick-off with a 30 minute introductory presentation that provides an overview of the purpose and goals of LASER Institute, including lessons learned from two virtual and one in-person cohorts of participating scholars.

Part 1 of the workshop will provide participant hands-on experience with the following types of instructional activities included in each curriculum module:

- **Interactive Presentations**. Each module contains slide decks for two interactive presentations: the first consisting of a conceptual overview of key terminology, techniques, and applications; the second presentation providing a short but highly structured code-along

activity that demonstrates key packages and functions required for specific data analysis techniques.
- **Coding Case Studies**. These interactive coding experiences can be completed by learners independently or in small groups and demonstrate key data-intensive research workflow processes (i.e., wrangling, visualizing, summarizing, modeling, and communicating data).
- **Readings and Discussion**. Essential readings are curated for participants to help them dive deeper into LA concepts, techniques, and applications introduced in presentation and case studies.
- **Software Tutorials**. Openly accessible software tutorials are curated for each module and are intended to help learners develop technical proficiency with essential software packages, functions, and programming syntax.
- **Badges & Microcredentials**. Each module includes a summative assessment activity designed to help learners reflect on how the concepts and techniques introduced in each lab might apply to their own STEM education research, where they can demonstrate their technical proficiency with the analytical techniques and methods addressed in each unit.

Part 2 of the workshop will consist of a presentation, facilitated discussion, and guided planning activities designed to support participants in setting up learning environments and adapting these materials for webinars, workshops, or courses at their home institution.

## 2.2 Recruitment

We anticipate having no difficulty in recruiting a diverse group of early and experienced scholars that are interested in incorporating LASER curriculum materials into their own teaching practice. Our recruitment strategy will involve both informal and formal approaches such as tapping into our existing professional networks and targeted digital marketing efforts on our established social media, e-mail, and web platforms. This built in audience includes key education stakeholders, researchers, educators, and current and past participants of the LASER Institute. To help offset the not insignificant expense of attending LAK this year, we will be using existing LASER Institute funding to offer substantial travel grants to our current and past LASER scholars that have already expressed an interest in developing local versions of LASER at their home institutions. Moreover, many of our current and past LASER scholars have expressed an interest in attending LAK as part of the professional learning plan they developed during LASER, so we will already have a large and highly motivated group likely to attend this workshop if offered.

## 2.3 Required Equipment

A projector and screen will be required by organizers, as well tables for collaboration. Attendees will need to bring laptops and will need adequate internet connectivity to participate in planned activities and access LASER curriculum materials. Specifically, participants will need to access to our freely available website that houses all the curriculum materials needed to teach, and learn from, the LASER curriculum. The website includes materials for each module including slide decks, videos, essential readings, discussion questions, case studies, tutorials, and assessment activities. The website will also include supporting materials for instructors such as pedagogical tips, information on computing infrastructure, technology stack, and logistics for set up. The source code for the LASER website and all instructional materials will be housed on the LASER Institute GitHub site, which allows for version control and collaborative editing of curriculum materials as well as the addition of new materials that may be developed by participants.

# 3 COMMUNICATION PLAN

## 3.1 Workshop Preparation

The workshop organizers will create a welcome packet to distribute to participants prior to the workshop. This packet will contain essential materials, including, but not limited to: information about the workshop facilitators; an overview and schedule for the day; a pre-workshop preparation checklist for setting up their LASER technology toolkit; links to websites that will be used throughout the day; and a planning template to support efforts for integrating materials post workshop.

## 3.2 Post Workshop Engagement

Participants will be invited to join the LASER community hub on Slack, which serves as a platform for building a community of practice and staying updated on future LASER events, resources and information. In addition, participants will be invited to participate in follow-up surveys designed to assess in what ways, and to what extent, participants have incorporated LASER curriculum materials into their own teaching and to gather feedback for improving these materials.

# REFERENCES

Alexander, B., Ashford-Rowe, K., Barajas-Murph, N., Dobbin, G., Knott, J., McCormack, M., ... & Weber, N. (2019). *Horizon report 2019 higher education edition* (pp. 3-41). EDU19.

Dede, C. (2016). Next steps for" big data" in education: Utilizing data-intensive research. *Educational Technology*, 37-42.

Krumm, A., Means, B., & Bienkowski, M. (2018). *Learning analytics goes to school: A collaborative approach to improving education*. Routledge.

Means, B., & Anderson, K. (2013). Expanding Evidence Approaches for Learning in a Digital World. *Office of Educational Technology, US Department of Education*.

Siemens, G. (2014). The journal of learning analytics: Supporting and promoting learning analytics research. *Journal of Learning Analytics*, *1*(1), 3-5.

# ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. DRL-2025090 and DRL-2321128 for the Learning Analytics in STEM Education Research (LASER) Institute. The authors would also like to thank the following members of the LASER team for their invaluable contributions: Co-PI Rob Moore and Senior Personnel Joshua Rosenberg, Jennifer Houchins and the support team members. The authors are also grateful to the over 120 scholars who have participated in this professional learning program and have helped to improve LASER through their active engagement and ongoing feedback.

# Fostering ethical and equitable learning analytics

**Rebecca Eynon[1], Simon Knight[2], Olga Viberg[3] and Alyssa Wise[4]**
[1]University of Oxford, UK
[2]University of Technology Sydney, Australia
[3]KTH Royal Institute of Technology, Sweden
[4]Vanderbilt University, US
edtechequity@education.ox.ac.uk

**ABSTRACT**: This submission is for a half-day workshop. There is a growing call for ethical and equitable learning analytics (LA), from educators, designers, researchers, and policy makers. Numerous frameworks and tool-based strategies have been produced for education. However, there has been limited focus on how these frameworks, tools and resources are identified, adopted and employed by those developing and implementing LA; the extent to which they find them valuable and meaningful in translating them into practice; the extent to which such frameworks and tools are effective in varied contexts and situations; what other kinds of resources are needed; and how the experiences of those doing this 'knowledge translation' are captured and disseminated in ways that can support the further development of responsible LA. This interactive workshop aims to address some of these challenges, through a combination of small group breakout sessions, guided discussion and the sharing of practice across the LA community.

**Keywords**: Equality, Ethics, Learning Analytics, AI, Responsible Learning Analytics, Knowledge Practices

## 1 BACKGROUND

The importance of the ethical use of data; tackling unintended bias and value judgements in the selection of data and algorithms; and the need to facilitate equity, fairness, and transparency in learning analytics in order to support positive social change in education systems is increasingly recognized by researchers, practitioners and policy makers that are part of the learning analytics (LA) community.

There is a growing and rich literature identifying the key issues and promoting a varied set of tool based and values based interventions and frameworks to support these commitments (Holmes, et al., 2022; Viberg et al., 2023). However, there is a need to further support the development of ethical and equitable learning analytics in practice (Baker & Hawn, 2022; Williamson & Kizilcec, 2022). This is a challenging area to navigate, as not only are there significant debates about the underlying philosophical position that these discussions involve (Hakimi et al., 2021;); but also, the recognition that often the attempt to encode complex social concepts, such as fairness, accountability, privacy, and equity into specific practices and guidelines is fraught with difficulty (Khalil, et al., 2023; Selwyn, 2019; Stark et al., 2021; Viberg et al. 2022).

In practice, this leads to two significant challenges for stakeholders working in this space. One is how to support the complex interchange of knowledge across varied "knowledge traditions" - as those in the LA community are tasked with translating work from varied related academic fields, such as, Philosophy and Technology, the Sociology of Education and Critical Data Studies, whilst also connecting with and designing with an awareness of policy and educator demands in a highly varied range of legal, cultural, educational, social, and technological contexts around the world (Eynon, 2023). Second is how to capture individual expertise and experience of those in the LA community who are developing responsible LA (Cerratto-Pargman et al., 2022) in a way that can be shared and further developed by others working in this space.

## 2 WORKSHOP OBJECTIVES AND OUTCOMES

The goal of this workshop is to address both of these knowledge translation challenges in order to support the development of ethical and equitable LA. It will focus on:

- Mapping the landscape of current resources available to the community to use in their practice
- Identifying the current dilemmas faced by stakeholders in developing responsible LA
- Sharing the varied ways that LA scholars are translating knowledge and expertise from different academic, practical and policy sources into their own practice; and the strengths and challenges of doing so
- Exploring the potential and need for other kinds of resources that reflect varied "real-world" experiences to support equitable and ethical LA
- Determining ways to better support dissemination of knowledge translation practices across the LA community.

Taken together, the workshop aims to develop a series of recommendations for how different stakeholders both within and beyond the LA community could support varied forms of knowledge creation and translation to inform the development of ethical and equitable learning analytics.

These recommendations will be written up as a short open access document synthesizing key outcomes and agreed follow-on activities. Such activities may include a special issue proposal on this topic, and future events to develop a LA community around these important issues. This will be hosted on the project website (see below) and promoted via social media.

## 3 WORKSHOP FORMAT

This interactive workshop will be held as a half-day event. The workshop will be hosted by four researchers from varied academic backgrounds who are working on different aspects of this challenge. Specific details have been removed from review but the workshop conveners' areas of focus include:

What kinds of resources / materials (inputs) are stakeholders using to develop responsible LA (outputs)? How can these resources, materials and impacts best be captured?

What are the practical ethical issues practitioners face when using LA, and what tools/resources have they found particularly helpful?

How can relevant guidelines (such as those from professional associations and journal editorial policies) foster ethics in LA in practice?

How can the 'real-world' experiences of educators and students of (in)equitable LA be distilled into usable resources for LA designers and practitioners?

Due to the discursive and interactive of the session, 15-20 participants from academia and practice will be recruited. This will be achieved via social media and the leveraging professional networks, supported by the four organizers being located in different countries and with varied disciplinary expertise.

Participants will have the opportunity to share their own approaches and experiences of this work and learn from others via a series of interactive sessions.

Interested participants will be asked to submit a 500-word extended abstract describing their perspective on how to develop ethical and/or equitable LA, drawing on their own experiences (e.g., communities engaged with, particular approaches and resources being used and how they are defining "ethical" and "equity" in their practice) and the strengths and challenges of their experiences to date.

### 3.1 Workshop Schedule

The draft workshop schedule is as follows:

1. Introductions (15 minutes)
2. 3x10 minute presentations by the workshop chairs on themes highlighted above (30 minutes)
3. 2 parallel break-out sessions and report back (60 minutes)
   Break-Out A: Resources for Responsible Analytics: Gap Analysis (brainstorming available resources to the LA community and relating the value of these for specific stakeholders and problems using a series of equity-based scenarios)
   Break-Out B: OER for Responsible Analytics (review and input into the development of open educational resources, based on 'real-world' experiences of educators, to help inform the development of equitable LA)

   Short Break (15 minutes)
4. Guided roundtable discussion: resources, dilemmas, practices and dissemination for responsible LA (75 minutes)
   Participants give 3-5-minute presentations, and the discussion will be facilitated by the workshop chairs, drawing on the themes from the submitted abstracts
5. Conclusions and next steps (15 minutes).

## 4 PLANNED MECHANISMS FOR COMMUNICATING

A webpage about the event will be created and hosted by the workshop conveners to advertise the event and encourage potential applications. This will include details about the intended content and structure of the event and the call for extended abstracts.

As above, this process will be further supported by the use of social media and conveners' professional Once participants have signed up to the event an email list will be created that will be used to communicate with the participants both prior and after the event. A shared google drive will enable the extended abstracts, alongside other resources and materials relevant to the workshop (both from the workshops conveners and the participants) to be shared. These will be maintained by the conveners, to facilitate further events and interactions beyond LAK 2024.

## REFERENCES


Baker, R.S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education,* 32, 1052–1092. https://doi.org/10.1007/s40593-021-00285-9

Cerratto Pargman, T., McGrath, C. et al., (2021). Responsible learning analytics: creating just, ethical, and caring. In: Companion Proceedings 11th International Conference on Learning Analytics & Knowledge (LAK21). https://oro.open.ac.uk/75925/

Eynon, R. (2023). The Future Trajectory of the AIED Community: Defining the 'Knowledge Tradition' in Critical Times. *International Journal of Artificial Intelligence in Education*, 1-6. https://doi.org/10.1007/s40593-023-00354-1

Hakimi, L., Eynon, R., & Murphy, V.A. (2021). The ethics of using digital trace data in education: A thematic review of the research landscape. *Review of Educational Research*, 91(5), 671-717. https://doi.org/10.3102/00346543211020116

Holmes, W., Porayska-Pomsta, K., Holstein, K. *et al.* (2022). Ethics of AI in Education: Towards a Community-Wide Framework. *International Journal of Artificial Intelligence in Education,* 32, 504–526. https://doi.org/10.1007/s40593-021-00239-1

Khalil, M., Prinsloo, P., & Slade, S. (2023). Fairness, Trust, Transparency, Equity, and Responsibility in Learning Analytics. *Journal of Learning Analytics*, 10(1), 1-7. https://doi.org/10.18608/jla.2023.7983

Selwyn, N. (2019). What's the problem with learning analytics? *Journal of Learning Analytics*, *6*(3), 11-19. https://doi.org/10.18608/jla.2019.63.3

Stark, L., Greene, D., & Hoffmann, A.L. (2021). Critical perspectives on governance mechanisms for AI/ML systems. *The cultural life of machine learning: An incursion into critical AI studies*, pp.257-280. https://doi.org/10.1007/978-3-030-56286-1

Viberg, O., Mutimukwe, C., & Grönlund, Åke. (2022). Privacy in LA Research: Understanding the Field to Improve the Practice. *Journal of Learning Analytics*, *9*(3), 169-182. https://doi.org/10.18608/jla.2022.7751

Viberg, O., Jivet, I., Scheffel, M. (2023). Designing Culturally Aware Learning Analytics: A Value Sensitive Perspective. In: Viberg, O., Grönlund, Å. (eds) Practicable Learning Analytics. Advances in Analytics for Learning and Teaching. Springer, Cham. https://doi.org/10.1007/978-3-031-27646-0_10

Williamson, K., & Kizilcec, R. (2022). Learning analytics dashboard research has neglected diversity, equity and inclusion. In, L@S' 21: Proceedings of the Eighth ACM Conference on Learning @ Scale, pp. 287-290, https://doi.org/10.1145/3430895.3460160

# Challenges and Opportunities of Learning Analytics Adoption in Higher Education Institutes: A European Perspective

L.W. van Meeuwen[1], B. Rienties[2], H. Drachsler[3,4], O. Viberg[5], R. Conijn[1], C.A. Vonk[1], J.W.A. Brijan[6], M. Specht[7]
[1] Eindhoven University of Technology, The Netherlands
[2] The Open University, The United Kingdom
[3] DIPF - Leibniz Institute for Research and Information in Education, Germany
[4] Open Universiteit, The Netherlands
[5] Royal Institute of Technology, Sweden
[6] University of Groningen, The Netherlands
[7] Delft University of Technology, The Netherlands
L.W.v.Meeuwen@TUe.nl

**ABSTRACT**: The successful application of Learning Analytics (LA) can improve student learning outcomes, student support and teaching. The key-challenges for LA adoption (i.e., Ethics, Leadership, Analytics culture, Analytics capabilities, Stakeholder involvement, and Technology) have been investigated. However, large-scale adoption remains lacking as does research into it. This half-day workshop organized in cooperation of several European universities has the aim to provide support to researchers and practitioners for realizing large scale adoption of practicable LA within higher education and the essential research thereon. Research and insights from the varied European contexts will be presented in this workshop for comparison with researchers from other non-European and other European contexts. The idea is that this exchange will provide insights to learn from the differences and overcome global challenges for successful adoption of LA at scale.

**Keywords**: Adoption at scale, Contextual challenges, Contextual opportunities, Higher education

## 1 INTRODUCTION

The practical application of Learning Analytics (LA) can benefit learning (e.g., Foster & Francis, 2020) and has the potential to further improve educational quality in higher education (Drachsler 2023; Praharaj, 2021; Seufert et al., 2019; Viberg & Gronlund, 2023). It is therefore valuable to see the ambition of institutions to put LA to practical use to improve educational quality (e.g., The Open University, 2015; Eindhoven University of Technology, 2018). A range of studies have shown that actual large-scale adoption by educators and institutions are still in its infancy more than a decade since the first Learning Analytics Knowledge conference was held in 2011 (Tsai et al. 2018, Hernandez-de-Menendez et al., 2022, Viberg et al., 2018) except for some examples (e.g., Herodotou et al., 2019). The key challenges for LA adoption include concerns around: Ethics, Leadership, Analytics culture, Analytics capabilities, Stakeholder involvement, and Technology (cf., Alzahrani, 2023). These challenges must be overcome for implementation at scale. Insights from different contexts can support this.

### 1.1 European Contexts

While the introduction of General Data Protection Regulation (GDPR; European Union, 2018) across Europe and the recent AI policy on data in teaching and learning for educators (European Commission,

2022) provide guidelines about how to deal with data protection and integration, it is widely documented that each member state has their own perspective on ethics, privacy and data (e.g., Drachsler & Greller, 2016; Korir et al., 2023). For example, German educational systems use a rather strict interpretation of data and ethics, while in the UK there seems to be a greater appetite for implementing learning analytics and data infrastructures to support students and educators. This might reflect the underlying cultural differences, interpretations of regulations, the engagement of key stakeholders, and ways of collaboration between institutions within a country (e.g., 4TU collaboration the Netherlands).

### 1.2 Comparing Contexts to Learn

Next to the differences within Europe, there are several aspects in the European context that bring up different challenges than in other international education systems, including GDPR and a more Humboldtian vision of higher education rather than a human capital perspective. In Europe, the barrier to obtaining data and getting started with LA in educational practice is often high. Institutions need to develop capabilities in ethics and privacy within a constantly changing environment (Knobbout et.al 2023; Prinsloo et al., 2023), which can be considered a data ecology rather than a closed ecosystem. Also, stakeholder engagement develops: conditions are being created and barriers overcome within institutions that allow early adopters to start pilot implementations in conjunction with research (e.g., Knobbout et al., 2023). To meet educational standards and adhere to institutional ethical and privacy guidelines, intricate design processes in collaboration with Educational Technology vendors are often requisite, as off-the-shelf solutions may not sufficiently address these criteria (Hernandez-de-Menendez et al., 2022; Knobbout et al., 2023; Tsai & Gašević, 2017; Drachsler & Greller, 2016). How do the two challenges listed below compare to other contexts in the European dimension, and what can stakeholders (like researchers) learn from other contexts to increase LA adoption globally?

### 1.3 Challenges

In summary, the two key challenges will be discussed in this workshop: Challenge 1: Although there is an increased uptake of LA applications that go beyond pilots (Leitner et al., 2017), few higher education institutions have yet implemented Learning Analytics at scale (Knobbout et al., 2023). Challenge 2: As a result, scientific output in this area of practical use of scalable LA in higher education is still limited (SoLAR, 2023).

## 2 THE WORKSHOP

More specifically, the aim of the workshop is to deal with the two challenges stated above by learning from other contexts. Therefore, the workshop will zoom in on contextual influences on two main bottlenecks for the issue of limited adoption at scale: 1. The process of obtaining data and 2. Stakeholder engagement (See Figure 1). From the research perspective, participants get to exchange expertise in the domain of practical application of LA from a different context (e.g., European, Asian, etc.); gain insights in research on practical application of LA (e.g., case studies); and get the opportunity to share state-of-the-art research on improving data processes and stakeholder engagement in other contexts. From a practical perspective, participants gain insight into potential bottlenecks for LA applications in their own institutions and simultaneously get tools to solve them.

The workshop design will allow for half a day meeting and comprises a combination of an Interactive Workshop and a Mini-track Symposium. The design includes discussions, group discussions, presentations, and voluntary contributions. We ask all participants to consider a research contribution detailing (case-)studies where LA benefits students learning, specifying approaches to improving data processes, and stakeholder engagement (such as participatory design and institutional collaboration). Contributions of 10-minute presentations may be submitted in the form of an abstract of up to 300 words. Based on the abstract, the workshop organization will carefully review the submissions and select a compelling array of diverse contributions that fit the workshop structure, workshop schedule, and enrich the workshop. Further interaction before and after the workshop will be supported by a website.

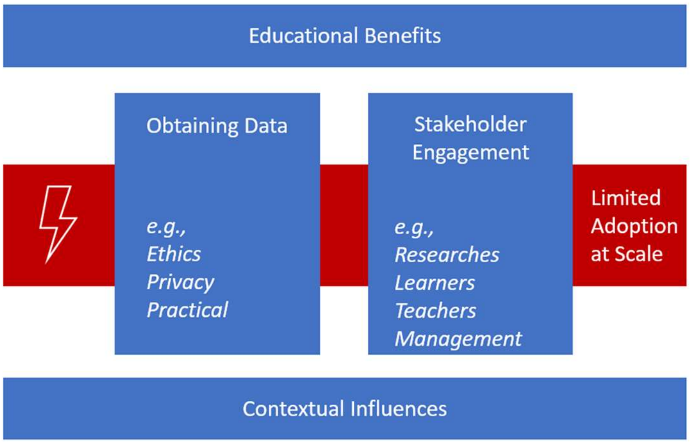



**Figure 1 Contextual influence on the issue of limited adoption of LA despite proven educational benefits. Workshop focuses on 1. the process of obtaining data and 2. stakeholder engagement**

The first organizer presents the scope-framework of the workshop during the opening. The workshop follows the two-pillar structure (see Figure 1). Each pillar will be introduced by the organization including examples from existing literature. The pillars will be enriched by the input from the participants who send in research contributions. Each pillar ends with group discussions on contextual impact of the bottlenecks to overcome and how and what we can learn from the differences. In the final part, the organization wraps up the workshop and summarizes the conclusions obtained in the break-out rooms. This will be used to sketch the open research questions and eye-openers on contextual influences that were obtained by comparing insights in LA adoption from several contexts.

## 2.1 Attract Participants

The organizing committee anticipates two groups of participants. First, participants who add a scientific contribution to the workshop (e.g., PhD candidates, Professors) and second, participants who want to share more practical experiences about contextual influences on LA adoption (e.g., (program)managers, research support staff) that can enable research. To approach both target groups, the organizers' professional networks will be used as well as the university alliance networks and the institutional members from SoLaR. In addition, (sample) posts will be made available for LinkedIn, and relevant mailing lists to recruit broadly. The workshop organization is aiming for 25 - 35 participants.

### 2.2 Conclusions

The findings and conclusions from the workshop will become available via the workshop website: https://sites.google.com/view/lak24adoptionworkshop/home

## REFERENCES


*Alzahrani, A. S., Tsai, Y. S., Iqbal, S., Marcos, P. M. M., Scheffel, M., Drachsler, H., ... & Gašević, D. (2023). Untangling connections between challenges in the adoption of learning analytics in higher education. Education and Information Technologies, 28(4), 4563-4595.*

*Drachsler, H., & Greller, W. (2016). Privacy and Analytics–it'sa DELICATE issue. A Checklist to establish trusted Learning Analytics. Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, Edinburgh.*

*Drachsler, H. (2023). Towards Highly Informative Learning Analytics. Open Universiteit, Netherlands, Heerlen. https://doi.org/10.25656/01:26787*

*Eindhoven University of Technology, 2023. TU/e Vision on Education. Retrieved from https://www.tue.nl/en/our-university/about-the-university/tue-strategy-2030/talent*

*European Commission. (2022). Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators. Publications Office of the European Union. https://doi.org/doi/10.2766/153756*

*European Union. (2018). General Data Protection Regulation. Retrieved from https://gdpr-info.eu/*

*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.*

*Herodotou, C., Rienties, B., Verdin, B., & Boroowa, A. (2019). Predictive learning analytics 'at scale': Guidelines to successful implementation in Higher Education based on the case of the Open University UK. Journal of learning Analytics, 6(1), 85-95.*

*Hernández-de-Menéndez, M., Morales-Menendez, R., Escobar, C. A., & Ramírez Mendoza, R. A. (2022). Learning analytics: state of the art. International Journal on Interactive Design and Manufacturing (IJIDeM), 16, 1209–1230. https://doi.org/10.1007/s12008-022-00930-0*

*Korir, M., Slade, S., Holmes, W., Héliot, Y., & Rienties, B. (2023). Investigating the dimensions of students' privacy concern in the collection, use and sharing of data for learning analytics. Computers in Human Behavior Reports, 9, 100262. https://doi.org/https://doi.org/10.1016/j.chbr.2022.100262*

*Knobbout, J., van der Stappen, E., Versendaal, J., & van de Wetering, R. (2023). Supporting Learning Analytics Adoption: Evaluating the Learning Analytics Capability Model in a Real-World Setting. Applied Sciences, 13(5), 3236.*

*Leitner, P., Khalil, M., & Ebner, M. (2017). Learning analytics in higher education—a literature review. Learning analytics: Fundaments, applications, and trends: A view of the current state of the art to enhance E-learning, 1-23.*

*Praharaj, S., Scheffel, M., Drachsler, H., & Specht, M. (2021). Literature review on co-located collaboration modeling using multimodal learning analytics—Can we go the whole nine yards?. IEEE Transactions on Learning Technologies, 14(3), 367-385.*

*Prinsloo, P., Khalil, M., & Slade, S. (2023, August). A Critical Consideration of the Ethical Implications in Learning Analytics as Data Ecology. In European Conference on Technology Enhanced Learning (pp. 371-382). Cham: Springer Nature Switzerland.*

*Seufert, S., Meier, C., Soellner, M., & Rietsche, R. (2019). A pedagogical perspective on big data and learning analytics: A conceptual model for digital learning support. Technology, Knowledge and Learning, 24, 599-619.*

*SoLAR (2023). SoLAR strategic plan April 2023 - May 2025. https://www.solaresearch.org/*

*The Open University (2015). Policy on ethical use of student data for learning analytics. The Open University. September 2014 (2015), 1–11.*

*Tsai, Y. S., & Gašević, D. (2017, March). Learning analytics in higher education---challenges and policies: a review of eight learning analytics policies. In Proceedings of the seventh international learning analytics & knowledge conference (pp. 233-242).*

*Tsai, Y. S., Moreno-Marcos, P. M., Tammets, K., Kollom, K., & Gašević, D. (2018, March). SHEILA policy framework: informing institutional strategies and policy processes of learning analytics. In Proceedings of the 8th international conference on learning analytics and knowledge (pp. 320-329).*

*Viberg, O., & Grönlund, Å. (Eds.). (2023). Practicable Learning Analytics. Springer Nature.*

*Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. Computers in Human Behavior, 89(December 2018), 98-110. https://doi.org/10.1016/j.chb.2018.07.027*

# 2nd Workshop on Highly Informative Learning Analytics (HILA)

**Lukas Menzel**
Goethe University Frankfurt, Germany
menzel@sd.uni-frankfurt.de

**Sebastian Gombert**
DIPF, Germany
s.gombert@dipf.de

**Tornike Giorgashvili**
Goethe University Frankfurt, Germany
giorgashvili@sd.uni-frankfurt.de

**Marcel Schmitz**
University of Applied Sciences Zuyd, Netherlands
marcel.schmitz@zuyd.nl

**Ioana Jivet**
FernUniversität in Hagen, Germany
Goethe University Frankfurt & DIPF, Germany
ioana.jivet@fernuni-hagen.de

**Hendrik Drachsler**
Goethe University Frankfurt & DIPF, Germany
drachsler@sd.uni-frankfurt.de

**ABSTRACT**: In the second iteration of the HILA workshop, we delve into the advanced frameworks of highly informative learning analytics, introducing a refined methodology for constructing data-enriched learning activities to provide highly informative feedback. This workshop employs an experiential, interactive format, enabling participants to assess the efficacy of the proposed methodology through real-world scenarios. The methodology is elucidated through the utilization of specialized methods and tools. These tools encompass FoLA2 design method for learning activities, Edutex as a learning analytics infrastructure, and various data-enriched learning activities (DeLA) for scientific writing, concept modelling, and collaboration that facilitate the gathering of comprehensive data that goes beyond standard log-data. The workshop concludes with an exploration of prospective use cases that align well with the introduced methodology.

**Keywords**: learning design, learning analytics indicators, psychometrics, feedback.

## 1 BACKGROUND

According to Hattie (2007), feedback has a powerful effect on learning success, with a mean effect size of d = 0.75. Wisniewski, Zierer, and Hattie (2020) even report a mean effect of d = 0.99 for highly informative feedback (on right/wrong, correct solution, type of processing, possibilities for improvement, hints on self-regulation, and learning strategies). Such feedback provides good conditions for self-directed learning (e.g., Winne & Hadwin, 2008) and effective metacognitive control of the learning process (Nelson & Narens, 1994). Until a few years ago, it was simply not possible in terms of personnel to provide highly informative and competence-oriented feedback at large university lectures. Nowadays, however, computers and other digital devices open up far-reaching possibilities that have not yet been fully exploited. This feedback has a high potential for improving individual study success and reducing dropouts, thus effectively supporting students in their learning process.

Within the 2nd iteration of the HILA workshop, we will work on an emerging design and development process for Highly Informative Learning Analytics based on various project experiences. We will first identify relevant Learning Analytics indicators for different learning activities. From there, we will demonstrate how we turned the designs of learning into data-enriched learning activities (DeLA) that have the potential to provide highly informative feedback. Finally, we discuss different types of feedback messages given to the students and future challenges for the HILA research.

## 2 PROPOSED SOLUTION

The workshop organisers have released several methodologies and tools for HILA. These are aimed at enhancing the quality of learning activities across diverse educational settings, ranging from K–12 to higher education. The workshop is modelled according to four phases, the first two quadrants in red and green (Identify & Combine phase) are addressed by the FoLA 2 methodology for collaboratively designing LA-powered learning activities ( Schmitz et al., 2022 ) and the OpenLAIR indicator repository (Ahmad et al., 2022). The third quadrant (Realise phase) is accomplished with the Edutex LA infrastructure (Ciordas-Hertel et al., 2021) and various DeLA created for most common learning activities in formal education settings, such as: 1. Reading-DeLA (Biedermann et al., 2023), 2. Writing-DeLA (Gombert et al., 2022), 3. Modeling-DeLA (Menzel et al., 2022), and 4. Collaborations-DeLA (Menzel et al., 2023). These applications can be used as instances of learning activities designed with FoLA 2 . These DeLAs generate a wealth of data within the Edutex infrastructure. Finally, the fourth quadrant (Research phase) is achieved by defining the process data indicators accordingly (Goldhammer et al., 2021; Drachsler, 2023).

The Fellowship of Learning Activities and Analytics (FoLA 2 ) is a methodology for designing learning activities with "analytics in mind". The method enables several participants with different roles to collaboratively interact with a set of card decks to create an LA-supported learning design. FoLA 2 can be used to develop, capture, and systematize design elements and to incorporate LA systematically. It takes advantage of the OpenLAIR indicator repository.

The Open Learning Analytics Indicator Repository (OpenLAIR) serves as an exhaustive compendium, encapsulating a decade-long evolution of indicators, metrics, and learning design activities within the specialized domain of learning analytics. In response to these findings, we have conceptualized a system designed to offer contextually relevant indicators and metrics, contingent upon the learning activities and events selected by educational stakeholders.

Edutex is a context-aware learning analytics infrastructure (Ciordas-Hertel, 2021). It can handle normal online learning activities and courses like those in Moodle, as well as physical learning environments equipped with various sensors.

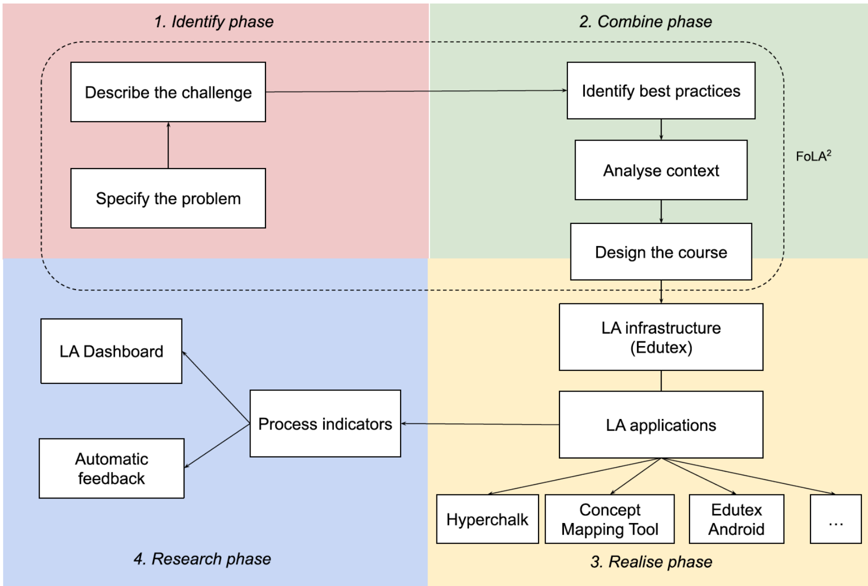



Figure 1: The *Highly Informative Learning Analytics* design process.

HyperChalk (Menzel et al., 2022) is a digital collaborative whiteboard built using the open-source component Excalidraw and a custom back-end. Similar to commercial whiteboard software such as Miro, Hyperchalk can be used to implement a wide range of creative collaboration tasks, but unlike commercial software, it allows researchers unlimited access to user data. It collects data appropriate for qualitative and quantitative studies on user behaviour demonstrated during collaboration tasks.

## 3 OBJECTIVES

In this workshop, we investigate the concept of highly informative learning analytics . The workshop is thought of as a hands-on, interactive session. We plan to demonstrate the proposed LA cycle in this workshop and allow the participants a hands-on experience. The workshop activities are divided as follows:

- Welcome and initial remarks
- A discussion of a representative task with the FoLA 2 methodology. The participants are divided into groups, each group is given a FoLA 2 board with which they need to design the learning session choosing among a set of available activities.
- The groups engage in a discussion in which they map the chosen activities to a set of existing tools.
- The participants explore the collaborative whiteboard tool Hyperchalk and the collaborative concept mapping tool.
- How to define the right process data indicators from the learning activities

## 4 ORGANISATION

The HILA workshop is organised as an interactive, full-day workshop. For the logistics, we need a large room (30 participants) with a reliable internet connection, projector, separate tables for group exercises, and, if possible, stationery such as sticky notes and pens. The organisers will provide technical tools and Slack channels and disseminate progress and outcomes via blogs and the Twitter hashtag #HILA24.

## REFERENCES


Ahmad, A., Schneider, J., Griffiths, D., Biedermann, D., Schiffner, D., Greller, W., and Drachsler, H. (2022). Connecting the dots – A literature review on learning analytics indicators from a learning design perspective. Journal of Computer Assisted Learning. https://doi.org/10.1111/jcal.12716

Ciordas-Hertel, G.-P.; Rödling, S.; Schneider, J.; Di Mitri, D.; Weidlich, J.; Drachsler, H. (2021). Mobile Sensing Mobile Sensing with Smart Wearables of the Physical Context of Distance Learning Students to Consider Its Effects on Learning. Sensors, 21, 6649. https://doi.org/10.3390/s21196649

Drachsler, H. (2023). Towards Highly Informative Learning Analytics. Open Universiteit, Netherlands, Heerlen. https://doi.org/10.25656/01:26787

Goldhammer, F., Hahnel, C., Kroehne, U., & Zehner, F. (2021). From byproduct to design factor: On validating the interpretation of process indicators based on log data. Large-Scale Assessments in Education , 9 (1), 20. https://doi.org/10.1186/s40536-021-00113-5

Gombert, S., Di Mitri, D., Karademir, O., Kubsch, M., Kolbe, H., Tautz, S., Grimm, A., Bohm, I., Neumann, K., & Drachsler, H. (2022). Coding energy knowledge in constructed responses with explainable NLP models. Journal of Computer Assisted Learning. 10.1111/ jcal.12767

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research , .77 (1), 16–17. https://doi.org/10.3102/003465430298487

Menzel, L., Gombert, S., Di Mitri, D., & Drachsler, H. (2022). Superpowers in the Classroom: Hyperchalk is an Online Whiteboard for Learning Analytics Data Collection. In I. Hilliger, P. J. Muñoz-Merino, T. De Laet, A. Ortega-Arranz, & T. Farrell (Eds.), Educating for a New Future: Making Sense of Technology-Enhanced Learning Adoption (pp. 463–469). Springer International Publishing. https://doi.org/10.1007/978-3-031-16290-9_37

Menzel, L., Gombert, S., Weidlich, J., Fink, A., Frey, A., Drachsler, H. (2023). Why You Should Give Your Students Automatic Process Feedback on Their Collaboration: Evidence from a Randomized Experiment. In: Viberg, O., Jivet, I., Muñoz-Merino, P., Perifanou, M., Papathoma, T. (eds) Responsive and Sustainable Educational Futures. EC-TEL 2023. Lecture Notes in Computer Science, vol 14200. Springer, Cham. https://doi.org/10.1007/978-3-031-42682-7_14

Nelson, T. O., & Narens, L. (1994). Why investigate Metacognition. In J. Metcalfe & A. P. Shimamura (Eds.), Metacognition. Knowing about Knowing. (pp. 1-25). Cambridge University Press.

Schmitz, M., Scheffel, M., Bemelmans, R., & Drachsler, H. (2022). FoLA2 — A Method for Co-creating Learning Analytics–Supported Learning Design. Journal of Learning Analytics, 9(2), 265-281. https://doi.org/10.18608/jla.2022.7643

Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. Frontiers in Psychology, 10. 1-14.

Winne, P. H., & Hadwin, A. F. (2008). The weave of motivation and self-regulated learning. In D. H. Schunk & B. J. Zimmerman (Eds.), Motivation and self-regulated learning: Theory, research, and applications (pp. 297–314). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

# Optimizing Human Learning. 4th Workshop eliciting Adaptive Sequences for Learning and Educational RecSys (WASL 2024)

**Jill-Jênn Vie[1], Yong Zheng[2], Fabrice Popineau[3], Yizhu Gao[4], Samuel Girard[1], Hisashi Kashima[5]**
[1] Inria, France
[2] Illinois Institute of Technology, Chicago, USA
[3] CentraleSupélec, France
[4] University of Georgia, Athens, GA, USA
[5] Kyoto University, Japan
jill-jenn.vie@inria.fr, yzheng66@iit.edu, fabrice.popineau@lisn.fr,
yizhu.gao@uga.edu, samuel.girard@inria.fr, kashima@i.kyoto-u.ac.jp

**ABSTRACT**: Personalized learning is an educational approach that tailors the learning experience to the unique needs, interests, abilities, and preferences of individual students. There are two major tools to assist personalized learning - adaptive learning systems (ALSs) that utilizes adaptive design, data-driven algorithms and personalized approaches to tailor learning experiences for individual students, and educational recommender systems (EdRec) which can deliver personalized recommendations (e.g., course, materials, pathway, peers) to users (e.g., students, educators and other educational stakeholders). These personalized learning technologies have proved remarkable success in enhancing user experiences. This rich full-day workshop, containing a keynote and tutorials, will target stakeholders and researchers about the background, motivations, knowledge, and skills related to ALSs and EdRec. It also provides an overview of emerging trends and unresolved challenges in these fields, delving into the technical challenges of framing reward functions: how learning analytics can be used to guide instruction, inform decision making for either teachers or intelligent tutoring systems. Furthermore, we aspire to foster a productive exchange of ideas with the audience and encourage more individuals to contribute their efforts towards advancing the area of personalized learning.

**Keywords**: Adaptive Learning, Personalized Learning, Reinforcement Learning, Recommender Systems, Bandits, Policy Learning, Causal Inference

## 1 ORGANIZATIONAL DETAILS

We propose an Interactive Workshop, full-day, containing hands-on tutorials about adaptive learning and educational recommender systems, and a keynote by Aritra Ghosh (Meta) and Andrew Lan (University of Massachussetts Amherst) about learning from sequential user data. To know more, check our website https://humanlearn.io.

**Audience:** The targeted audience is anyone who is interested in personalized learning and recommender systems. To cater to the audience members without prior experience or knowledge, our tutorial will include an introductory talk about them (e.g., the background and knowledge of recommender systems), along with more in-depth discussions about our topics, and technical details of the algorithms and implementations for the advanced users.

**Program committee:** Hisashi Kashima, Fabrice Popineau, Jill-Jênn Vie, Jacob Whitehill (WPI).

## 2 WORKSHOP OBJECTIVES

Technology-Enhanced Learning (TEL) leverages various technological tools and innovations to enhance the learning experience, making it more engaging, personalized, and effective for learners of all ages and backgrounds (Balacheff et al., 2009; Kirkwood et al., 2014). Personalized learning plays a vital role in TEL, and it was advanced by two technologies. One of them is the adaptive learning systems (ALSs) that emerge as a pioneering and promising TEL field (Bingham et al., 2018; Wang et al., 2023). ALSs enhance the scalability and resourcefulness of personalized education, a feat that would be difficult to accomplish in conventional classroom environments. Educational recommender systems (EdRec), as another TEL technology, have been utilized in education to improve personalized learning experiences, help fill students' knowledge gaps, recommend appropriate learning materials (formal and informal materials), courses, pathways and after-school programs, and adapt learning to context-aware or mobile environments, and so forth (Garcia-Martinez et al., 2013; Drachsler et al., 2015; Erdt et al., 2015; Khribi et al., 2015; Klašnja-Milićević et al., 2015; Zheng, 2023). These platforms collect a massive amount of data over various profiles, that can be used to improve learning experience: intelligent tutoring systems can infer what activities worked for different types of students in the past, and apply this knowledge to instruct new students. In order to learn effectively and efficiently, the experience should be adaptive: the sequence of activities should be tailored to the abilities and needs of each learner, in order to keep them stimulated and avoid boredom, confusion and dropout.

Implementing ALSs brings various pedagogical benefits, including accelerated learning, remediation, immediate feedback, and interactive learning (Hattie, 2008). To maximize these benefits, researchers from both academia and industry attempt to develop new systems by incorporating cutting-edge techniques (e.g., conversational AI, reinforcement learning from human feedback). As efforts to develop and implement ALSs accumulate, it becomes increasingly evident that the effectiveness in promoting learning achievement varies across different systems. While substantial progress has been made in constructing ALSs, a noticeable gap persists in our understanding of the specific architectural and design choices that contribute to their efficacy (Imhof et al., 2020; Muñoz et al., 2022). This workshop seeks to address this knowledge gap starting with an overview of the historical development and architecture of ALS and adaptive approaches, followed by a detailed analysis of the potential impact factors influencing the effectiveness of ALSs in facilitating learning achievement. In the context of reinforcement learning, we want to learn a policy to administer exercises or resources to individual students (Bassen et al., 2020; Clement et al. 2015; Lan et al. 2016; Whitehill et al., 2017). Still, framing the reward is key for making these tools effective. We will discuss future directions and persistent challenges that ALSs and LAK community must address to propel this field forward (Motz et al., 2023).

Recommender systems (RecSys) have seen widespread adoption across various internet applications. These applications encompass e-commerce platforms like Amazon.com, online streaming services such as YouTube, and social media platforms like Facebook. The remarkable success of these applications in enhancing user experiences and aiding decision-making through personalized recommendations underscores the effectiveness of RecSys. Recently, RecSys has witnessed significant progress for optimizing interaction, click-through-rate, or profit, thanks to a range of intriguing and promising areas like multi-task learning (Zhang et al., 2023), multi-objective optimization (Zheng et al., 2022), multi-stakeholder considerations (Zheng et al., 2021), and addressing issues related to fairness, accountability, and transparency (Shin et al., 2019; Deldjoo et al., 2023), among others. However, these advancements in the realm of RecSys have not been sufficiently shared with the education community or in the development of EdRec. Can we use similar methods to enhance the performance of teaching in order to promote lifetime success? When optimizing human learning, which metrics should be optimized (Doroudi et al., 2019)? Learner progress? Learner retention? User addiction? The diversity or coverage of the proposed activities? This workshop aims to bridge this gap by providing the audience with an in-depth understanding of the background, motivations, and the essential knowledge and skills required for EdRec. Additionally, it will offer a concise overview of emerging topics and the unresolved challenges currently shaping this field. Student modeling for optimizing human learning is a rich and complex task that gathers methods from machine learning, cognitive science, educational data mining and psychometrics (Bergner et al., 2018).

## REFERENCES

Balacheff, N., Ludvigsen, S., De Jong, T., Lazonder, A., Barnes, S. A., & Montandon, L. (2009). Technology-enhanced learning. Berlin: Springer.

Bassen, Jonathan, et al. "Reinforcement learning for the adaptive scheduling of educational activities." Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020.

Bergner, Yoav, Geraldine Gray, and Charles Lang. "What does methodology mean for learning analytics?." Journal of Learning Analytics 5.2 (2018): 1-8.

Bingham, A. J., Pane, J. F., Steiner, E. D., & Hamilton, L. S. (2018). Ahead of the curve: Implementation challenges in personalized learning school models. Educational Policy, 32(3), 454–489.

Clement, Benjamin, et al. "Multi-Armed Bandits for Intelligent Tutoring Systems." Journal of Educational Data Mining 7.2 (2015): 20-48.

Deldjoo, Y., Jannach, D., Bellogin, A., Difonzo, A., & Zanzonelli, D. (2023). Fairness in recommender systems: research landscape and future directions. User Modeling and User-Adapted Interaction, 1-50.

Doroudi, Shayan, Vincent Aleven, and Emma Brunskill. "Where's the reward? a review of reinforcement learning for instructional sequencing." International Journal of Artificial Intelligence in Education 29 (2019): 568-620.

Drachsler, H., Verbert, K., Santos, O. C., & Manouselis, N. (2015). Panorama of recommender systems to support learning. Recommender systems handbook, 421-451.

Erdt, M., Fernandez, A., & Rensing, C. (2015). Evaluating recommender systems for technology enhanced learning: a quantitative survey. IEEE Transactions on Learning Technologies, 8(4), 326-344.

Garcia-Martinez, S., & Hamou-Lhadj, A. (2013). Educational recommender systems: A pedagogical-focused perspective. Multimedia services in intelligent environments: Recommendation services, 113-124.

Hattie, J. (2008). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. Abingdon: Routledge.

Imhof, C., Bergamin, P., & McGarrity, S. (2020). Implementation of adaptive learning systems: Current state and potential. Online teaching and learning in higher education, 93–115.

Khribi, M. K., Jemni, M., & Nasraoui, O. (2015). Recommendation systems for personalized technology-enhanced learning. Ubiquitous learning environments and technologies, 159-180.

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.

Klašnja-Milićević, A., Ivanović, M., & Nanopoulos, A. (2015). Recommender systems in e-learning environments: a survey of the state-of-the-art and possible extensions. Artificial Intelligence Review, 44, 571-604.

Lan, Andrew S., and Richard G. Baraniuk. "A Contextual Bandits Framework for Personalized Learning Action Selection." on Educational Data Mining: 424. 2016.

Motz, Benjamin A., Yoav Bergner, Christopher A. Brooks, Anna Gladden, Geraldine Gray, Charles Lang, Warren Li, Fernando Marmolejo-Ramos, and Joshua D. Quick. "A LAK of Direction Misalignment Between the Goals of Learning Analytics and its Research Scholarship." (2023).

Muñoz, J. L. R., Ojeda, F. M., Jurado, D. L. A., Peña, P. F. P., Carranza, C. P. M., Berríos, H. Q., ... & Vasquez-Pauca, M. J. (2022). Systematic review of adaptive learning technology for learning in higher education. Eurasian Journal of Educational Research, 98(98), 221–233.

Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. Computers in Human Behavior, 98, 277-284.

Wang, S., Christensen, C., Cui, W., Tong, R., Yarnall, L., Shear, L., & Feng, M. (2023). When adaptive learning is effective learning: comparison of an adaptive learning system to teacher-led instruction. Interactive Learning Environments, 31(2), 793-803.

Weidlich, Joshua, Dragan Gaševic, and Hendrik Drachsler. "Causal Inference and Bias in Learning Analytics: A Primer on Pitfalls Using Directed Acyclic Graphs." Journal of Learning Analytics 9.3 (2022): 183-199.

Whitehill, Jacob, and Javier Movellan. "Approximately optimal teaching of approximately optimal learners." IEEE Transactions on Learning Technologies 11.2 (2017): 152-164.

Zhang, M., Yin, R., Yang, Z., Wang, Y., & Li, K. (2023). Advances and Challenges of Multi-task Learning Method in Recommender System: A Survey. arXiv preprint arXiv:2305.13843.

Zheng, Y. (2023). Tutorial: Educational Recommender Systems. In International Conference on Artificial Intelligence in Education (pp. 50-56). Cham: Springer Nature Switzerland.

Zheng, Y., & Wang, D. X. (2022). A survey of recommender systems with multi-objective optimization. Neurocomputing, 474, 141-153.

Zheng, Y., & Toribio, J. R. (2021). The role of transparency in multi-stakeholder educational recommendations. User modeling and user-adapted interaction, 31, 513-540.

# Culture and Values in Learning Analytics: A Human-Centered Design and Research Approach

**Luis P. Prieto**
Universidad de Valladolid (Spain)
luispablo.prieto@uva.es

**Olga Viberg**
KTH Stockholm (Sweden)
oviberg@kth.se

**María Jesús Rodríguez-Triana**
Tallinn University (Estonia)
mjrt@tlu.ee

**Ioana Jivet**
FernUniversität in Hagen (Germany)
ioana.jivet@fernuni-hagen.de

**Bodong Chen**
University of Pennsylvania (US)
cbd@upenn.edu

**Maren Scheffel**
Ruhr-University Bochum (Germany)
maren.scheffel@rub.de

**ABSTRACT**: The increasing complexity and potential for disruption of recent artificial intelligence (AI) advances (including their applications to education, like learning analytics – LA) make the issue of aligning those technologies with human stakeholders' goals and values more relevant than ever. Further, the global reach of many of these educational applications puts into question whether LA designed for one setting and its local culture can be unproblematically transferred to another culture/setting. The fields of cross-cultural psychology and human-computer interaction have developed frameworks and methods to explicitly and systematically look at such problems of value alignment in the design of technology (e.g., under the label "value-sensitive design"), but there are still few examples of applying them in the LA domain. In the present workshop, participants will learn about and apply different value-sensitive design methods and value frameworks, to the design (or re-design) of specific LA innovations proposed by the facilitators and participants themselves.

**Keywords**: values, culture, value-sensitive design, learning analytics, human-centered design.

## 1 INTRODUCTION

The recent eruption of artificial intelligence (AI) into the public discourse -in line with this year's conference theme "Learning Analytics in the Age of Artificial Intelligence"- has prompted many questions about its impact on education (including issues of assessment, accountability, and literacies). These new questions, along with long-standing issues like fairness and bias in machine learning and AI as well as other ethics and privacy issues are pushing the learning analytics (LA) community to call for a more humane alignment of the LA systems we design (e.g., the notion of human-centered LA, see Buckingham Shum et al., 2019). However, in LA research and practice we have not systematically and explicitly taken such goals and values into account, with only a few examples available (Campos et al., 2023; Chen & Zhu, 2019). Luckily, we can turn to other fields that have been working on frameworks and methods to both model and investigate the goals, needs and values of groups and individuals, at different levels.

From the field of cross-cultural psychology, it is now well-established that culture is a primary way in which certain values are reflected. One of the ways to examine and understand culture is through its values. Cultural values are understood as “collective tendencies to prefer a certain course of events above another, expressed by qualifications such as good and bad, dirty and clear, ugly and beautiful” (Hofstede et al., 2010). According to Viberg et al. (2023), the values emphasized in a society may be “the most central feature of culture” (Schwartz, 2006, p. 139) as these values describe a shared understanding of what society views as good, right and desirable (Williams, 1970). For example, if a society values success and ambition, this might be reflected in “a highly competitive economic system [...] and child-rearing practices that pressure children to achieve” (Schwartz, 2006, p. 139). In an educational setting, such an environment might foster competition among students as ‘being better than your peers’ defines a successful learner, encouraging the use of social comparison features in the design of LA dashboards (Jivet et al., 2017). It is also well established in psychology that values, i.e., what a person considers important in life, are also a key motivational construct at the individual level, related to well-being, planned behavior (including learning behavior) and even neural correlates (Sagiv & Schwartz, 2022).

In the field of human-computer interaction, building on this rich social sciences research background, value-sensitive design (VSD) has been proposed as “a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner” (Friedman et al., 2017). Yet, VSD is more than a philosophy, and it has developed specific methods to consider human values (e.g., privacy, trust, and autonomy) in a systematic fashion throughout system design and research processes. Thus, VSD holds great potential as a concrete way to consider cultural aspects and individual values in LA (Chen & Zhu, 2019; Viberg et al., 2023). The main expected benefits of using VSD in LA include: making LA more relevant to a wider range of stakeholders and facilitating (and understanding) the transfer of LA innovations to new contexts and across cultures.

This workshop aims to both popularize VSD in learning analytics and help participants incorporate such methods (and cultural and individual value considerations more generally) into LA design and research processes. The workshop builds upon two previously separate workshop series at the LAK and EC-TEL conferences (workshop names blinded for review), thus combining the efforts, contents, and experiences for both of those aspects to provide workshop participants with the best of both.

## 2 WORKSHOP GOALS AND STRUCTURE

Against this backdrop, the present workshop aims to:

1. introduce the participants to cultural considerations and VSD methods;

2. further explore and raise awareness of possible influences of stakeholders' values and preferences on the acceptance of LA systems, the design of LA tools, and the evaluations of LA interventions;

3. practice selected VSD methods that can be used to inform more responsible and human-centered design of LA and AI in education; and

4. invite participants to jointly plan how to incorporate VSD methods into actual LA design or research processes, e.g., to perform transfer or multi-site/multi-cultural LA studies.

To achieve these aims, the workshop took the form of a "design challenge" (where participants work in small teams to de-construct and re-design specific LA systems, or to plan cross-cultural LA studies), with non-expositional scaffolding from the organizers.

## ACKNOWLEDGEMENTS

This workshop is supported by grant RYC2021-032273-I, financed by MCIN/AEI/ 10.13039/501100011033 and the European Union's NextGenerationEU/PRTR. It has also received support from the Regional Government of Castile and Leon, under project grant VA176P23.

## REFERENCES

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

Campos, F., Nguyen, H., Ahn, J., & Jackson, K. (2023). Leveraging cultural forms in human-centred learning analytics design. *British Journal of Educational Technology*, bjet.13384. https://doi.org/10.1111/bjet.13384

Chen, B., & Zhu, H. (2019). Towards Value-Sensitive Learning Analytics Design. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 343–352. https://doi.org/10.1145/3303772.3303798

Friedman, B., Hendry, D. G., Borning, A., & others. (2017). A survey of value sensitive design methods. *Foundations and Trends® in Human–Computer Interaction*, *11*(2), 63–125.

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind: Intercultural cooperation and its importance for survival*. McGraw-Hill.

Jivet, I., Scheffel, M., Drachsler, H., & Specht, M. (2017). Awareness is not enough: Pitfalls of learning analytics dashboards in the educational practice. *European Conference on Technology Enhanced Learning*, 82–96.

Sagiv, L., & Schwartz, S. H. (2022). Personal Values Across Cultures. *Annual Review of Psychology*, *73*(1), 517–546. https://doi.org/10.1146/annurev-psych-020821-125100

Schwartz, S. (2006). A Theory of Cultural Value Orientations: Explication and Applications. *Comparative Sociology*, *5*(2–3), 137–182. https://doi.org/10.1163/156913306778667357

Viberg, O., Jivet, I., & Scheffel, M. (2023). Designing Culturally Aware Learning Analytics: A Value Sensitive Perspective. In O. Viberg & Å. Grönlund (Eds.), *Practicable Learning Analytics* (pp. 177–192). Springer International Publishing. https://doi.org/10.1007/978-3-031-27646-0_10

Williams, R. M. (1970). *American society: A sociological interpretation*. Knopf.